
THE VALUE OF CERTAIN MEASUREMENTS IN THE TRAINING OF TEACHERS

EXPERIMENTALLY DETERMINED

This book is presented

by

The Government of the United States

as an expression of

Friendship and Goodwill

of the

People of the United States

towards

The People of India

THE JOHNS HOPKINS UNIVERSITY
STUDIES IN EDUCATION

No. 9

EDITED BY
EDWARD F. BUCHNER

THE VALUE OF CERTAIN MEASUREMENTS IN THE TRAINING OF TEACHERS

EXPERIMENTALLY DETERMINED

AGNES SN ..., Ph. D.

BALTIMORE
THE JOHNS HOPKINS PRESS
1928

The Lord Baltimore Press
BALTIMORE, MD., U. S. A.

EDITOR'S NOTE

The development of objective tests in education has been passing through the various critical stages involved in rendering them more and more perfect instruments of procedures. The methods of refinement relied upon have been closely related to the inner technique of mathematical theory. During a period of intensive use of these tests, the author of this monograph came upon the problem of the possible value of an early introduction to an intelligent use of tests and scales for determining a more effective procedure in the training of teachers. In March, 1924, Dr. Snyder received the first award of the Fellowship for Research in Education offered by Pi Lambda Theta, the honorary fraternity for women in education. Its generous stipend enabled her to carry through a larger project during 1924-1925 with the hope of finding a more definite answer to her question: Does skill in using selected educational tests function in teacher-training? The Department of Education of The Johns Hopkins University, in appreciation of the award, welcomed the unusual opportunity thus offered to hunt for any new values accruing in the use of tests. It seemed reasonable to hope that all past evidences of the value of the use of standardized measures by teachers in service would be enhanced by discovering some positive values in the course of teacher-training. While it is recognized that the author's findings are not absolutely conclusive, it is hoped that her efforts to analyze a complicated situation will point the way for new experimental efforts to measure changes in the early growth in teaching skills and attitudes by means of resultant changes in pupil learning.

E. F. B.

ACKNOWLEDGMENTS

The present study was made financially possible through the award of a research fellowship to the author by the Pi Lambda Theta Fraternity of Women in Education. Grateful appreciation is here expressed to Pi Lambda Theta.

Indebtedness is gratefully acknowledged to Dr. Edward F. Buchner, Dr. Florence E. Bamberger, and Dr. Fowler D. Brooks, of The Johns Hopkins University, who gave most generously of their time and their interest in the planning and the criticism of the study. Indebtedness is due and appreciation expressed to Miss Lida Lee Tall, Principal of The Maryland State Normal School at Towson, who placed the resources of the school at the service of the experiment and whose constant encouragement and support did much to bring the work to completion. At the same institution, acknowledgment of indebtedness is likewise made to Mr. Norman Woelfel, Instructor in Tests and Measurements, who by incorporating the scoring of much of the test material in his courses assisted materially; to Miss Anna D. Halberg, Miss Stella Brown, Mr. Allan Hulsizer, and Miss Alice O'Neill, who cooperated in making the experiment part of the regular work of the students in the training centers under their supervision; to Miss Gertrude Carley, registrar, who gave valuable assistance in the collection of the data; to the faculty and student body who assisted in the scoring. Grateful acknowledgment is expressed to Dr. John L. Stenquist, Director of the Baltimore Bureau of Statistics and Research, who gave valuable help through his support of the experiment and through placing the facilities of the bureau at the service of the experiment.

Appreciation is expressed to Dr. Henry J. West, Superintendent of Public Schools of Baltimore, and to Mr. Clarence G. Cooper, Superintendent of Schools of Baltimore County, for permission to conduct the experiment in the schools under their administration. Appreciation of their

hearty cooperation is expressed to the principals of the schools in which the experiment was conducted: Mr. David G. Butterfield, School 99; Mr. Martin M. Hihn, School 83; Miss Ella L. Smith, School 212; Mr. Marshall Stitely, School 22; and Miss Martha Stromberg, School 86.

Indebtedness is due and acknowledged to Dr. Edward L. Thorndike for supplying copies of the I. E. R. Intelligence Tests used in testing the senior class at The Maryland State Normal School at Towson. Likewise, indebtedness is due and acknowledged to the Baltimore City school administration for supplying the tests used in Baltimore City. In the course of the experiment, the author had occasion to send questionnaires to many educators throughout the country. For their careful replies, grateful acknowledgment is made.

Appreciation is expressed to the teachers of practice whose unfailing cooperation in the direct supervision of the experimental student teaching formed the core of the experiment: Miss Winifred E. Barrett, Miss Mary L. Broening, Miss Anna M. Chesno, Miss Helen M. Lerian, Miss Marguerite K. Stotler, and Miss Mildred E. Tyson, of Baltimore City; Miss Gertrude Buckley, Miss Mildred Buckley, Miss Margaret Ewing, and Miss Katharine Logan, of Baltimore County; Miss Ruth E. Buckley, Miss Sadie Fitzgerald, and Miss Caroline E. Reed, of the Campus Elementary School at Towson. Appreciation is expressed to the teachers of the control classes in Baltimore County who permitted time to be used in testing their classes, although the nature of the experiment prevented any returns from the testing program to be made to them: Miss Margaret Feeney, Miss Margaret Grau, and Miss Mary L. Hipsley.

Finally, appreciation is expressed to the students who participated in the experiment. Their conscientious effort in carrying out directions and in making accurate reports was of the utmost value in this work.

AGNES SNYDER.

TABLE OF CONTENTS

CHAPTER I

The Present Status of Educational Measurements in Teacher Training Institutions

I. SURVEY OF THE LITERATURE

The field of educational literature is rich in studies dealing with the practical application of the measurement movement to classroom procedure. Nothing, however, has been done to subject the value of tests and measurements in the pre-service training of teachers to experimental study. It is true that the journals report occasional articles which indicate that tests and measurements are being used in the training of teachers; but, as a rule, such studies are conducted on a small scale and with a view to the immediate improvement of instruction rather than with the purpose of putting the worth of the procedure in the training of student teachers to experimental test. Of such type is a report by Emma B. Grant.[1] Here, a class of 34 senior students in a city training school tested and diagnosed the reading ability of a practice class and made definite recommendations for follow-up work to the teacher of practice. M. E. James[2] gives a similar report of work done in the State Normal School at Salem, Massachusetts. Mary A. Grupe and Elsa M. Smith[3] give evidence that work of the children in the fundamentals is kept up to standard in training classes. There is implication that the student teachers participate in the work, but there is no indication that emphasis is placed upon the definite training of the students in the use of the results of testing programs.

[1] Grant, "Motivating the Course in Tests and Measurements for the Teacher-in-Training," *Journal of Educational Method*, IV, 1925.

[2] James, "Using the Results of Measurement in Reading in Training Student Teachers," *Elementary School Journal*, XXII, 1922, pp. 190-96.

[3] Grupe and Smith, "The Use of Educational Measurements in the Training Department of the State Normal School, Ellensburg, Washington," *Educational Administration*, VII, 1921, pp. 517-26.

Because the yield of reported studies was so small and because also it was believed desirable to collect certain contributory data, it was decided to use the questionnaire method in order to determine the present status of the use of tests and measurements in teacher training institutions. The questionnaires were differentiated to suit the three main types of institutions dealing with teacher training:

1. Those with no other function than teacher training; i. e., teacher training and normal schools.
2. Departments or schools of education in colleges and universities.
3. Bureaus of Research in their cooperative function in teacher training.

A summary only of the results of this phase of the work is given here. A complete report is on file in the library of The Johns Hopkins University.

2. EDUCATIONAL MEASUREMENTS IN TEACHER TRAINING INSTITUTIONS

Questionnaires (Appendix A) were sent in October, 1924, to the 279 public teacher training institutions listed in Bulletin, 1924, No. 10, Department of the Interior, Bureau of Education, *Statistics of Teachers Colleges and Normal Schools 1921-22.* Replies were received as follows: Alabama, 3; California, 3; Colorado, 2; Connecticut, 1; Georgia, 1; Hawaii, 1; Illinois, 3; Indiana, 3; Iowa, 2; Louisiana, 2; Maine, 1; Maryland, 4; Massachusetts, 3; Michigan, 15; Minnesota, 3; Missouri, 3; Montana, 1; New Hampshire, 2; New Jersey, 2; New York, 9; North Carolina, 4; North Dakota, 4; Ohio, 11; Oklahoma, 4; South Dakota, 1; Texas, 1; Virginia, 1. In all, 90 blanks were returned either wholly or partially answered and, in addition, 5 letters were received which gave information of value to the study. The replies included 56 state, 1 territorial, 15 city, and 23 county institutions. The following types of institutions were represented: 1 School of Education, 4 Teacher Training Schools, 6 City Normal Schools, 3 Teachers Colleges, 1 City Normal Training School, 3 County Training Schools, 20 County Normal Schools, 34 State Normal Schools, 22 State Teachers Colleges, Normal Colleges and Normal Universities, and 1 Territorial Normal Training School.

A summary of the results is offered in order of the nine headings of the questionnaire:

1. Catalogue Name or Names of Courses in Which Training in the Use of Educational Tests and Measurements is Given.—From 95 replies, we learn that 68 institutions gave definite courses in tests and measurements, 18 included such instruction in other courses, and 9—for the most part small county normal schools—gave no such courses. The relative frequency of the three practices—the specific course, inclusion in other courses, and no course—is indicated in terms of percent in Figure 1. The institutions giving specific courses

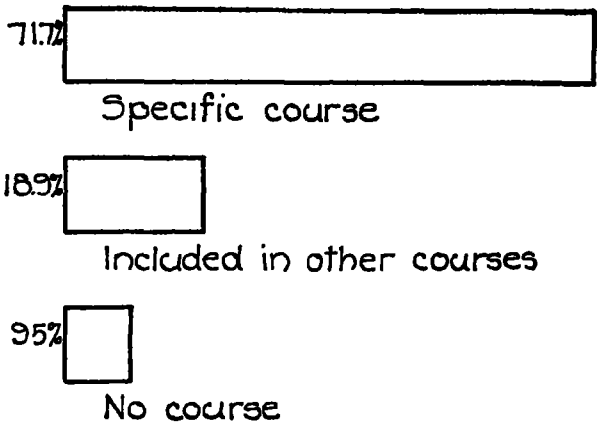

FIGURE 1.—The relative frequency in 95 teacher training institutions of those giving specific courses in tests and measurements, of those including such work in other courses, and of those giving no such courses

in tests and measurements listed such courses variously, but, in general, the designations fell into the following classification:

Courses including one or both words, "tests" and "measurements," in title	46
Courses including one or both words, "mental" and "intelligence," in title	7
Courses giving psychological designations	9
Courses giving only catalogue numbers	18
Total	80

It will be noted that the 68 institutions gave a total of 80 courses, indicating an average of 1.18 courses to an institution.

2. Enrollment.—There were 42 definite replies. It was found that in 26 institutions in which courses in tests and measurements were required or in which some of such courses were required and some elective, there was an average total enrollment of 302.5 and an average enrollment in courses in tests and measurements of 62.5. In 16 institutions having elective courses only, the average total enrollment was 969 with an average enrollment in courses in tests and measurements of 30. Thus it is seen that the institutions in which courses in tests and measurements were elective were the larger institutions, while the enrollment in these elective courses was very much smaller than in institutions in which the courses were required.

3. Teaching Staff.—There were 46 inclusive replies. Of these, only 5 indicated that full-time instructors in tests and measurements were employed, while 41 indicated that part of one or more instructors' time was so employed. The median number on the entire teaching staff was 36, while the median amount of time given to instruction in tests and measurements was one-fifth of one instructor's time.

4. School Year.—The 59 replies indicated a school year ranging from 32 to 48 weeks with both median and Q_1 at 38 weeks. The length of courses in tests and measurements ranged from 6 to 36 weeks with both Q_1 and the median at 12 and Q_3 at 18. The trimester division of the school year prevailed, and courses in tests and measurements were, in general, limited to one 12-week term.

5. Salaries.—There were so many different interpretations of this item that, although it was treated in nearly all of the blanks, only 44 answers could be used. The data indicate that where full time or more than full time of one instructor was given to courses in tests and measurements, the average expenditure was $3000, but that the prevailing condition, i. e., the use of part of an instructor's time for such courses,

required, on the average, an annual expenditure of $806.50 for one such course covering a period of 12 weeks.

6. Expenditure Other Than Salaries.—There were but 24 answers sufficiently inclusive to be of value. Again, the expenditure was slight, amounting to a median expenditure of $50 for equipment and supplies and $26.67 for assistance. The meagre amounts given may mean, as some letters accompanying the questionnaires seemed to indicate, that the budget system employed made the accurate accounting of specific items such as this, practically impossible.

7. Testing Program in Schools Outside of Institution.—There were but 33 complete replies to this section indicating that a testing program was conducted as part of a course in tests and measurements. In addition, 16 stated definitely that no such program was carried on, while 9 others stated that there was such a program but gave no details. The results show an evident effort to make courses in tests and measurements practical by incorporating testing programs in them. As a rule, such programs were limited to a small number of schools, frequently to the training school connected with the institution.

8. Cost of Testing Program.—There were but 17 complete replies. These indicated expenditures ranging from $5 to $700 with a median expenditure of $50 which was defrayed in 10 out of the 17 cases by the institution, 3 by the institution and the cooperating school system, and 4 by other sources.

9. Use Other Than the More Effective Training of Students Made of the Data Gathered in the Testing Program.—There were 39 replies to this question. Of these, 12 indicated that no use other than the more effective instruction of the students was made of the data. The 27 replies indicating that some other use was made of the results may be summarized as follows:

	Replies
For classification of pupils	9
Supervision of instruction	14
Use in superintendent's office	1
Instruction and classification	2
Special experimental work	1

3. EDUCATIONAL MEASUREMENTS IN DEPARTMENTS AND SCHOOLS OF EDUCATION IN COLLEGES AND UNIVERSITIES

Scope of the Investigation.—Questionnaires (Appendix A) were sent to the state university of each state and to 10 leading universities. In all, 24 replies, representative of 21 states and one island territory, were received. The following states and territory were represented: California, Colorado, Florida, Illinois, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Nevada, New Hampshire, New York, North Carolina, Rhode Island, South Dakota, Tennessee, Washington, Wisconsin, and Porto Rico. The situation thus revealed may be summarized as follows:

1. Catalogue Name or Names of Courses in Which Training in the Use of Educational Tests and Measurements is Given.—There were 22 replies to this item with a total of 53 courses or an average of 2.41 courses to an institution. In 12 cases, catalogue numbers only were given to designate the course, while in the 41 remaining institutions, 30 different titles were indicated. Because of the variety of designations and the implication of differentiation in the field of tests and measurements which such variety gives, these titles are given in full:

Educational Tests and Measurements	5
Tests and Measurements	3
Educational Measurements	3
Statistical Methods in Education	3
Intelligence and Educational Measurements	1
Tests of Intelligence	1
Administration of Education	1
Mental and Educational Measurements	2
Individual and Group Intelligence Testing	1
Administration of Tests, and Tabulations, Interpretation, and Use of Test Results	1
Seminary of Educational and Intelligence Testing	1
Educational Psychology	1
Educational Diagnosis	1
Statistical Method	1
Mental Tests and Mental Diagnosis	1
Individual Mental Examinations	1
Psycho-Educational Clinic	1
Measurement and Experiment in Elementary Education	1
Measurement in High Schools	1
Measurement in Secondary Education	1

Measurement and Experiment in Kindergarten and First Grade	1
Educational Measurement in Elementary School	1
Measurement in Secondary Schools	1
Problems in Educational Measurement	1
Experimental Education	1
Personnel Research in High Schools	1
Personnel Research in Elementary School	1
Introduction to Educational Measurements	1
Seminary in Educational Measurements	1
Elementary School Surveys and Tests	1
Total	41

2. Enrollment.—There were 14 replies to this item. Enrollment in entire departments ranged from 122 to 2500 with a median enrollment of 268.5, while enrollment in courses in tests and measurements ranged from 6 to 1473 with a median enrollment of 43. In 9 cases, the courses were designated as elective, in 2 as required, and in 1 as required for special curricula. It will be noted that this condition is the opposite of that prevailing in teacher training institutions where, in by far the greater number of cases, courses in tests and measurements were required.

3. Teaching Staff.—There were but 18 complete answers to this unit of which but 2 indicate that instructors gave full time to courses in tests and measurements. In those institutions in which part of one or more instructors' time was used for such courses, the amount of time ranged from 1/20 to $2\frac{1}{2}$ with a median of $\frac{1}{4}$ of one instructor's time. Here, the condition was found to be similar to that in teacher training institutions, the prevailing tendency being toward the use of one instructor's time, the average fractional part of such time being somewhat greater than in teacher training institutions.

4. School Year.—There were 22 answers to this item. They indicated a school year ranging from 32 to 48 weeks with median, Q_1, and Q_3 at 36, and courses in tests and measurements ranging from 12 to 40 weeks with median at 18, Q_1 at 17, and Q_3 at 34 weeks. The semester course was shown to be the prevailing one in length of time as was the trimester course in teacher training institutions.

5. Salaries.—The 15 replies indicated a range of expenditures on salaries of entire departments of education of from

$8400 to $147,825 with a median expenditure of $19,400 and a range of expenditures on salaries of instructors in tests and measurements of from $500 to $8850 with a median expenditure of $1610.

6. Expenditure Other Than Salaries.—There were 14 replies. These indicated a range of expenditure for equipment and supplies of from $17 to $500 with a median expenditure of $75 and a range of expenditure on assistance of from $0 to $2000 with a median of $0. These amounts are not very different from those similarly spent by teacher training institutions, the amount on equipment and supplies being slightly greater and that on assistance being slightly less in the present instance.

7. Testing Program in Schools Outside of Institution.—There were 15 replies indicating that a testing program had been carried out in connection with the course or courses and 3 indicating that there had been no such program. Again, as in the case of teacher training institutions, effort has evidently been made to make courses in tests and measurements practical by incorporating testing programs in them.

8. Cost of Testing Program.—There were but 5 replies indicating the cost of testing programs. These ranged from $0 to $2250 with a median of $300. Of 11 replies indicating how the expenses were defrayed, there were 3 in which such expenses were born entirely by the institution, 3 by the institution and other sources, and 5 by other sources, the last including local schools, the system surveyed, the schools tested, and workers in the class. The replies indicated a somewhat higher expenditure than in teacher training institutions and that expenses were borne either in entirety or in part by outside agencies to a greater extent than in teacher training institutions.

9. Use Other Than the More Effective Training of Students Made of the Data Gathered in the Testing Program.—Since there were but 13 replies, these are quoted here:

"Used by supervisors of schools." "Effort was made to discover special aptitudes or difficulties." "No." "Yes, study of Junior High School achievement in the light of educational and vocational choices." "Yes, to help the school." "Each school supplied with individual

record cards, class records, general summaries for system, recommendations for reclassification and remedial instruction." "Used by principals in classifying students; by teachers in evaluating work; by professor in charge to illustrate measurement in public talks." "Classification of pupils; basis for supervision and remedial teaching, scaling of tests; sundry studies of educational problems, etc." "Results given to local teachers." "Classification use in the schools tested." "Usual administrative uses. Each teacher, as a rule, examined thoroughly her own pupils and used results for sectioning, grading, promoting, etc." "Classification into sections of a junior high school." "By local principal and teachers in work of supervision."

4. COOPERATION OF BUREAUS OF EDUCATIONAL RESEARCH WITH TEACHER TRAINING INSTITUTIONS

Questionnaires (Appendix B) were sent to the 64 Research Bureaus maintained by city systems and to the 8 of the 49 research agencies maintained by higher institutions not included in institutions to whom one or the other of the 2 questionnaires had been sent. Of this total of 72, 65 replies were received. It is interesting to note that 35 different designations appeared in the titles of these bureaus. Of these, the most common was "Department of Educational Research," the designation given to 20 of the number. The next most common was "Department of Research," the name given to 6. Other names indicated specific functions such as "guidance," "statistics," "classification," "building program," "psychological research," "measurement," "efficiency," "survey," "attendance," "child study." The following summary of results is offered:

1. Cooperation with Teacher Training Institution or Department or School of Education of a College or University.—Of the 65 total replies, 44 indicated some kind of cooperation. Of these, 39 indicated the type institution with which cooperation was carried on:

State Universities	26
State Teachers' Colleges	9
Normal Schools	3
City Training Schools	1
Total	39

2. Courses Given in Educational Institutions by Members of Research Bureau.—There were 65 replies. Of these, 30 indicated that members of the research staff gave courses in

tests and measurements in cooperating institutions, while 35 indicated that no such instruction was given. Of the 30 giving such instruction, 5 specified that the instruction was given in extension courses, and 1 that a course was given to teachers and principals.

3. Length of Courses in Weeks Per Year; Number of Periods Per Week; Length of Periods.—There were 35 complete replies. The number of weeks per year ranged from 6 to 42 with a median of 18; the number of periods per week ranged from 1 to 7 with a median of 2, and the length of the periods ranged from 50 minutes to 2 hours with a median of 1 hour.

4. How the Cost of Instruction Is Met.—There were 29 answers as follows:

Cost met entirely by teacher training institution	26
Cost met by bureau of research and educational institution cooperatively	3
Total	29

5. Amount Appropriated in 1923-24 by the Bureau of Research for Instructional Purposes.—But one of the three institutions which indicated that the instructional cost was met by the bureau of research in cooperation with the educational institution answered Item 5. In this case the amount was $2200.

6. Testing Programs in Cooperation with the Teacher Training Institution.—There were 33 replies to this question, 22 of which indicated that cooperative testing programs were conducted and 11 that no such cooperative testing programs were conducted.

7 and 8. Amount Spent on Testing Programs (Both Items Include Details of Costs).—There were but 8 complete replies to these items, these 8 having a range in total expenditures of from $10 to $5150 with a median expenditure of $740. There were 11 replies indicating a range of from $50 to $3000 with a median of $500 expended on clerical assistance.

9. Benefits of Cooperative Plan to Teacher Training Institution and to Bureau of Research.—There were 23 replies.

The benefits to the teacher training institution indicated were: The advantage of close contact with the field; the coordination of theory and practice; the opportunity of extending the work of the institution to teachers in service; and the supply of current data for research. The benefits to the bureau indicated were: The greater interest of the teachers in the program of the bureau; the advantage of securing student help in carrying out testing programs; and the extension of the service of the bureau.

5. COMPARISON OF RESULTS OF STUDY OF THE USE OF TESTS AND MEASUREMENTS BY THE THREE TYPES OF INSTITUTIONS INVESTIGATED

The preceding investigation is based upon a study of 95 teacher training institutions, 24 departments or schools of education in colleges and universities and 62 Bureaus of Educational Research, giving a total of 184 institutions. The replies indicated the following conditions:

1. Courses in tests and measurements were general in teacher training institutions and departments or schools of education in colleges and universities. There was also a tendency to utilize the services of members of the staffs of Bureaus of Research in the giving of such courses in cooperation with teacher training institutions which were, in the majority of cases, departments or schools of education in colleges and universities. To a lesser extent, more so in teacher training institutions than in colleges and universities, instruction in the use of tests and measurements was given in other educational courses; e. g., " School Management " and " Principles of Education."

2. Courses in tests and measurements were for the most part required in teacher training institutions and elective in colleges and universities.

3. In both types of institutions—and also in the courses given by members of the staffs of Bureaus of Research—the instructors in these courses gave only part of their time to this field of work.

4. While the inadequacy of the data made a thorough-going study of costs impossible, the expenditure for all phases of this work appeared to be low.

5. Testing programs generally accompanied such courses, effort usually being made to give service to the schools tested while providing professional education in the use of tests to the students in the course.

6. The tendency of having the schools or systems tested share in the expense of the testing program was indicated particularly in the case of colleges and universities.

7. While various expressions of the use made of such testing programs were given, in no case was there indication that they were used as part of a definite program in the training of student teachers as is the case in the present study.

CHAPTER II

THE PROBLEM, THE CRITERIA, AND THE PLAN OF SOLUTION

I. THE PROBLEM

The preceding chapter shows a general tendency to include courses in tests and measurements in the curricula of teacher training institutions. It also shows an effort to make such courses practical by linking them with testing programs and with instructional procedure in elementary school situations. The present study is an attempt to measure the worth of instruction in the use of tests and measurements in so far as such instruction enters into the direct pre-service training of teachers during the student teaching period. The study makes no attempt to measure the general value of courses in tests and measurements as part of a teacher training curriculum. Neither does it attempt to measure the value of the apparently prevalent practice of having the students in such courses participate in test surveys of school systems. Instead, it confines itself to what is assumed should be the point of emphasis in such courses, i. e., instruction in the use of the results of standard tests in teaching the children tested.

With the above limitation in mind and with the further limitations of scope necessitated by the practical considerations of time and available school situations, the following statement of the problem is given:

The place and value of training student teachers to use skillfully the results from general and diagnostic tests in the instruction of children.

Will a typical group of seniors in a normal school grow more in ability to teach children if they base their instruction upon the analyses of general and diagnostic tests than will a similar group of seniors who do not use tests as a basis of instruction?

Both groups teach for equal lengths of time in Grades 3, 4, and 5, the first group spending three thirty-minute periods per week in test-determined teaching of reading; an equal

amount in arithmetic (fundamentals and problems); and three fifteen-minute periods per week in spelling over a period of not less than six weeks.

2. THE CRITERIA

Since the problem is centered about the effect of the use of the results of tests upon the growth in teaching skill of the student teacher, the criteria selected for the determination of such growth in skill are of the utmost importance. Study of the literature on the rating of teaching shows a general tendency to place such rating upon a twofold basis:

1. Progress made by the children taught.
2. The worth of instruction as observable in the classroom.

It is generally accepted that if sufficiently accurate and adequate measures of pupils' progress could be secured, practically no other criterion need be considered. E. C. Elliott, for example, urges that stress in teacher rating be placed upon "clearly-defined results of teaching rather than upon formal procedure or upon personal factors of presumptive educational worth."[1]

The difficulty of getting such a measure of results is clearly expressed by W. M. Davidson as follows: "The effect of the teacher's qualities upon the pupils may be outstanding; but they may be so subtle as to escape hasty inexpert inspection. Effects upon the literary taste of the pupils or upon their attitude toward problems of conduct, for example, are far more subtle than effects upon their accuracy in spelling or upon their facility in arithmetic. . . . . It sometimes happens that the supervisor tends to ascribe to the qualities of the teacher effects which are simply due to the increasing maturity and out-of-school experience of the pupils. The qualities observable in the teacher should serve as a check upon this fallacy."[2]

[1] Elliott, E. C. "How Shall the Merit of Teachers be Tested and Recorded?" *Educational Administration and Supervision*, I, 1915, pp. 291-299.

[2] Davidson, W. M. "How to Measure the Efficiency of Teachers," *National Education Association Addresses and Proceedings*, 1913, p. 286.

Again, S. A. Courtis, in mapping out a program for placing teacher rating upon a scientific basis makes "definition of teaching ability wholly in terms of the changes to be produced in children" the first step. He expresses, however, the difficulty of this step and the inadequacy of our present measures in the following statement: "Self-development, self-appraisal, self-control, cooperation do not show in standard tests. That is why, at present, there is so little correlation between teaching ability and scores in standard tests."[1]

B. F. Pittinger gives three possible planes upon which a teacher can be rated:

1. The plane of results or of pupil achievement.
2. The plane of the teaching and learning process.
3. The plane of the teacher's equipment for teaching, both native and acquired.[2]

He would measure the plane of results only, were it possible to do so effectively; but because of the impossibility of measuring all the results of teaching or of picking out from the body of measured results any single teacher's contribution, he recommends concentration upon the classroom process as being more measurable, the rating to include both teacher and pupil activities.

Compared with the general problem of measuring teacher growth, the problem presented in this study is simple. Here, the immediate interest is limited to the teaching of three subjects, progress in which can be submitted to objective test. Hence *a comparison between the growth of the children of the experimental and control classes in reading, spelling, and arithmetic abilities as measured by the difference between the scores made in the standard tests in those subjects given at the beginning and those made in the tests given at the end of the experiment is used as the main criterion* of the growth in teaching skill of the two groups of students teaching those classes.

[1] Courtis, S. A. "Standards of Teaching Ability," *Educational Review*, LXII, p. 185.
[2] Pittinger, B. F. "Problems of Teacher Measurement," *Journal of Educational Psychology*, VIII, p. 104.

While it is understood that the above criterion would serve to establish the growth in teaching skill only in the three subjects tested, even within these narrow limits other factors, as suggested by N. N. Davidson, might operate to obscure the importance of the training of the student teacher to use test results as a factor in the progress made by the children. In order to determine as accurately as possible the relative influence of such training as measured by the progress of the children, it was decided to measure, in so far as the circumstances surrounding the present experiment permitted, every possible variable that might affect the progress of the children.

If, measured by this criterion, the present study should indicate that better teachers of reading, spelling, and arithmetic could be produced through training in the use of test results, or even if the indications should point to the results in but one of these three subjects, the results would be of significance in the general problem of teacher training. If, on the other hand, the results should indicate that such specific training affected the general teaching skill of the student teacher, the significance of such procedure in the teacher training problem would be immeasurably greater. To get such a general measure means the use of the second criterion mentioned above, i. e., the worth of instruction as observable in the classroom.

In the present status of the teacher rating problem, a measure of the worth of instruction as observable in the classroom is so clouded in obscurities as to render any results obtained of doubtful validity. In the full realization of the inadequacy of such a measure as a criterion, the desirability of having it seemed of such advantage in the present study that it was decided to make an attempt to secure it. Hence, a second criterion is used; i. e., *a comparison between the growth in teaching skill of the student teachers of the experimental and control classes as measured by the difference between their ratings at the beginning and at the end of their student teaching periods.*

Again, as in the case of the growth in the ability of the children, the effect of the training or the absence of training in

the use of standard test results in their teaching may be obscured by the operation of other factors. Again, therefore, as in the case of the factors operating to influence the progress of the children, a measure of the relative influence of every such factor which the circumstances surrounding the experiment rendered possible of measurement, was secured.

It is important that the differences between the two criteria used be kept in mind. The first, the growth of the student teacher as measured by the growth of the children, is specific and objective. The second, the growth in teaching skill of the student teacher as observable in the classroom, is general and subjective. These differences establish, from the point of view of validity, the relative importance of the two criteria. The first is, of course, the more valid, and, hence, more weight will be given to the significance of the results as measured by it.

3. THE PLAN

In brief, the solution of the problem required:

1. Equated experimental and control groups of children and student teachers.

2. Batteries of standard tests given to the children, one form at the beginning and another form at the end of the experiment.

3. Rating of the student teachers at the beginning and at the end of the experiment.

4. Comparison of the growth made by the two groups of children, experimental and control, as judged by the difference between the results of the first and second batteries of tests for the two groups.

5. Comparison of the growth made by the two groups of students, experimental and control, as judged by the difference between the two ratings of their teaching.

6. Measurement of all possible variables that may have affected the results of the experiment, such as the mental and chronological ages of the children and of the student teachers.

CHAPTER III

The Conditions

1. THE EXPERIMENTAL SITUATION

The experiment was conducted in twenty-three practice classes connected with The Maryland State Normal School at Towson. Three distinct types of schools were involved: the city graded school, the rural one-or-two room school, and the elementary school on the normal school campus. There were 12 classes in Baltimore City located in 5 different schools in 5 different sections of the city; 8 classes in Baltimore County located in 6 one-or-two-room schools in 6 different sections of Baltimore County; 3 classes in the elementary school on the normal school campus. In the city, the 12 classes consisted of 4 each of Grades 3, 4, and 5. In the county, there were 2 third-grade groups and 6 combined 4th and 5th grade groups. In the campus school there was 1 class each of Grades 3, 4, and 5.

While the 5 city schools used are located in different sections of Baltimore, the conditions are similar in so far as the population of each of the neighborhoods of these schools consists, for the most part, of typical city workers. In no case is there an appreciable foreign element, nor are the extremes of economic background represented. The 6 schools in Baltimore County are in neighborhoods of small farms, the population consisting for the most part of farmers, but including, especially in the case of one school neighborhood, mill workers and mechanics. The campus school population represents the better type of suburban life, with rather more favorable economic background than is found in the other schools represented.

In the city schools, two classes of the same grade were in charge of one teacher of practice who divided her time between the two classes, teaching part of the time but spending

the greater part of the day in supervision of student teachers.[1] In the county schools and in the campus school, the teacher of practice has but one classroom in charge. In each situation—city, county and campus—there is a director of practice, in the case of the campus school the function of the director of practice being vested in the principal of the school.

The total numbers of pupils in these classes were as follows:

	Pupils
12 city classes	490
8 county classes	170
3 campus classes	83
Total	743

Apportionment of Time Used in Experiment.—The Maryland State Normal School at Towson is organized on a trimester basis, each term consisting of 12 weeks. The experiment was so planned that the first term was used for giving the necessary tests to the children in the practice classes and to the students involved and for studying the results of this testing. The experimental teaching was conducted during the 2nd and 3rd terms, thus involving two different groups of student teachers.

The Testing Program: Elementary Pupils.—The testing program of the children included 7 achievement tests and 2 intelligence tests.

The tests given in October and November, 1924, were as follows:

Intelligence:
- Mentimeter School Group 2A
- Illinois Examination I Form 1

Achievement:

Reading
- Thorndike McCall Reading Scale, Form 1
- Stanford Achievement Test Form A: Test 3—Word Meaning; Test 2—Sentence Meaning; Test 1—Paragraph Meaning
- Monroe's Standardized Silent Reading Test Revised, Test 1, Form 1

Spelling
- Morrison McCall Spelling Test, Form 1
- Stanford Achievement, Form A, Test 6 (Primary) or Test 9 (Advanced)

[1] This condition has since been changed, the teacher of practice in the city now being responsible for but one room as in the rural practice classes.

Arithmetic Problem Solution
  The Buckingham Scale for Problems in Arithmetic, Form 1
  The Stanford Achievement Test Form A, Test 5
Arithmetic Calculation
  Monroe Survey Scale in Arithmetic 1
  Stanford Achievement Test Form A, Test 4
  Courtis Cleveland Survey

In the final testing, the intelligence tests were omitted. Otherwise, the achievement tests were repeated in other forms with the exception of the Courtis Cleveland Survey which exists in but one form. Form 4 was used of the Thorndike McCall Reading and the Morrison McCall Spelling. In other cases, the second form of the test was used.

To assure uniform results, the experimenter attempted to give all of the fall tests herself. This meant spreading the testing program over a period of two months. It was found necessary toward the end of the program to use the assistance of two trained workers besides in order to prevent invalidating the results through an unduly prolonged testing period. Care was taken to give all of one type test within a week in order to minimize the difference in the results obtained due to spread of time. The scoring of these tests was made a part of the regular work in the courses in *Tests and Measurements* but again the amount of work involved was beyond the possibility of such treatment and other assistance had to be employed. In either case, the tests were scored twice, the experimenter making the final check wherever discrepancies occurred.

In order to carry the experimental teaching as late into the spring as possible, it was found necessary to abandon the original plan of having all of the testing done by the experimenter. Instead, the experimenter instructed a member of the faculty in each city school—in each case the member selected having had courses in tests and measurements and other experience in testing—in giving the specific tests involved. This faculty member then gave all of the tests in the classes used in the experiment in her school. This plan insured disinterested testers, for none had had any contact with the experiment. The tests in the county schools and in the campus school were given by the instructor in tests and measurements

at the Normal School, his assistant, the two directors of practice, and a trained member of the Normal School staff.

The Testing Program: Normal School Students.—One test, the I. E. R. Intelligence Test, was given to the students. The giving of this test occupied two full mornings, or approximately seven hours. The experimenter instructed members of the staff experienced in testing work in the giving of the test. These instructors gave the tests to the fifteen sections comprising the senior class, each part of the test being given at the same time and, in so far as it is possible to do so where different personalities are involved, in the same way. The scoring was done for the most part by volunteer faculty and student help. It was not possible to score every test twice as was done with the tests given to the children. Instead, a random sampling of each set of tests was checked and in cases where errors were found, the entire set was rescored.

2. GENERAL STATUS OF THE PUPILS: MENTAL, CHRONOLOGICAL, AND EDUCATIONAL AGES

Although through equating the groups, through change of classes, through leaving school and through other causes, the numbers in the course of the experiment were greatly reduced, and, therefore, the status of the groups as reported in this chapter does not represent the status of the groups which persisted throughout the experiment, the general status of the groups as indicated by mental, chronological, and educational ages will, nevertheless, be reported here. This is done because, although no measure of its influence is as yet available, it seems reasonable to assume that the intelligence and achievement of the group of which an individual is a part does help determine his success. The data are here offered in the hope that in them some further hypotheses might be found for pursuing further the search for the causes of school progress in which broad field the present study is laid.

Attempt has been made in Figure 2 to bring together the educational, the mental, and the chronological ages of the pupils in the three types of schools—campus, city and rural—so that the relationships among these three ages and the like-

nesses and differences among their relationships in these schools might be clearly seen. The average mental age was computed by averaging the mental ages of the Trabue Mentimeter and of the Illinois Intelligence Examination for each pupil and then averaging these averages by grades. The educational ages are the equivalents of the scores made on the Stanford Achievement Test. In order to establish a basis for comparison, the upper and lower limits of normal chronological ages at entering Grades 3, 4, and 5, according to the Strayer-Englehardt Computation Table, are shown in Figure 2. The averages for the mental, educational ,and chronological ages for the three grades of the three types of schools are as follows:

	Campus	Rural	City
Grade 3	M. A. 111.8 mos. C. A. 107.2 E. A. 111.8 (25 pupils)	M. A. 99.5 mos. C. A. 112.5 E. A. 108.9 (23 pupils)	M. A. 93.05 mos. C. A. 106.4 E. A. 106.5 (159 pupils)
Grade 4	M. A. 126.3 mos. C. A. 115.8 E. A. 126.5 (29 pupils)	M. A. 117.8 mos. C. A. 126.4 E. A. 124.7 (86 pupils)	M. A. 112.3 mos. C. A. 116.8 E. A. 120.2 (175 pupils)
Grade 5	M. A. 148.1 mos. C. A. 122.5 E. A. 137.9 (29 pupils)	M. A. 130.9 mos. C. A. 143.5 E. A. 137.2 (61 pupils)	M. A. 130.6 mos. C. A. 133.8 E. A. 134.3 (156 pupils)

The three groups of schools present three different conditions. In the campus school, the average chronological age for each grade lies about the upper normal age limit for entering the grade, while both the average mental and educational ages exceed the chronological age. Here, then, is a school situation in which the pupils, on the average, have superior capacity and in the third and fourth grades are achieving in accordance with that capacity, though in the fifth grade the achievement, while still high as compared with chronological age, is not so high as the capacity.

In the rural schools, the average chronological age in each grade exceeds the upper limit of the normal chronological age for entering the grade. The average mental age, on the contrary, lies about the upper limits of the normal chronological

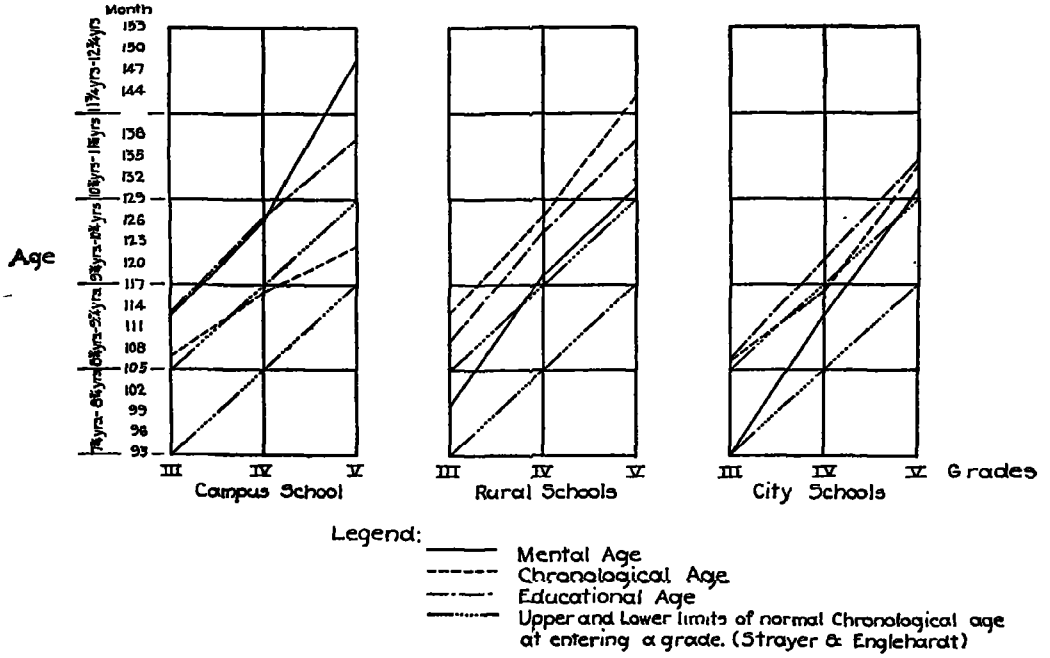

FIGURE 2.—Average mental age, educational age and chronological age of 83 campus, 170 rural and 490 city pupils. Total 743 in grades 3, 4, and 5.

age for entering the grade. The educational age exceeds the mental age but does not reach the chronological age levels. Here, then, is a situation where achievement exceeds capacity, but where neither capacity nor achievement reaches the chronological age averages.

In the city schools, the average chronological age corresponds to the upper normal chronological age limit for entering the grades. The average mental age is lower than the average chronological age but lies within the normal chronological age limits for the grades. The average educational age exceeds both the average mental and chronological ages. Here, then, capacity levels are below chronological levels, but achievement levels are higher than either.

In comparing the three groups, it is seen that the average chronological age of the third and fourth grades campus and city schools are approximately the same but that the fifth grade average of the city group is considerably higher than the fifth grade average in the campus school. The average chronological age in the rural schools is appreciably higher in each grade than in either of the other groups.

In mental age, the averages in the third and fourth grades are lower in the city school than in either the rural or the campus schools while the fifth grade city and rural are approximately the same. The average mental age in the campus school is as appreciably higher than it is in the other two groups of schools as is the average chronological age lower.

While there are these diversities among the three groups in average chronological and mental ages, there is no corresponding diversity in average educational age. In other words, the three groups of schools differ but little in average achievement but have reached such achievement status at different average ages and with different average mental capacity. The children in the campus school have reached the achievement indicated at a comparatively early age with high mental equipment; the children in the rural schools, at a much later age with less mental equipment reach approximately the same achievement status, a status which, however, is still not equal to their average chronological age; the pupils of the city

schools at an average chronological age less than that of the pupils of the rural schools, with a correspondingly lower average mental age, reach approximately the same average achievement age, an age higher than the average chronological age.

While these data are presented in terms of averages only and, therefore, give but a rough estimate of the relative conditions in the three situations, the widely different conditions in these fundamental factors should be kept in mind in interpreting the results of the present experiment. If growth, for instance, has proceeded at such different rates as these suggest it is reasonable to expect different results in the three situations in the present experiment.

3. GENERAL STATUS OF THE STUDENT TEACHERS AS INDICATED BY THEIR PREVIOUS TRAINING, SEX, INTELLIGENCE RATING, AND CHRONOLOGICAL AGE

Numbers Participating in the Experiment.—In all, 158 students of the senior class, which numbered 494, participated in the experiment. As stated above, the experimental teaching was confined to the second and third terms of the school year 1924-25. The 158 students participating were distributed as follows:

Second Term:		Third Term:	
City Classes	42	City Classes	36
Rural Classes	20	Rural Classes	23
Campus Classes	17	Campus Classes	20
	79		79

In the middle of each term, the rural and campus school students were shifted in order to give them at least two experiences during their student teaching, the students who began their student teaching in the campus school going to the rural centers and those beginning in the rural classes going to the campus school. Such shift was also made among the city students during the second term but it was not made during the third term. The numbers given above are for the beginning of each term. Some losses, incurred during each term, will be accounted for in Chapter IV.

The previous training of the students varied as follows:

Of the 80 students in the rural group (those teaching in rural and campus classes):
74 attended 39 different high schools within the state, the greatest number attending any one high school being 7.
3 attended private institutions within the state.
3 attended institutions without the state.
Of the 78 students in the city group:
47 attended one of the two girls' high schools in the city.
21 attended the other of the two girls' high schools in the city.
6 attended other preparatory schools within the state.
4 attended institutions without the state.

Of the 158 students, 155 were women and 3 were men, a somewhat smaller proportion of men than in the entire class.

The I. E. R. Intelligence Examination was given to the senior class at the Maryland State Normal School at Towson on November 19 and 20, 1924. The complete distribution as well as the distribution of the students participating in the experiment is shown in Figure 3. The intelligence records of 3 of these students were lacking; hence the distribution is given for 155 students instead of the complete number, 158. It will be noted that the distribution of the scores of the group in the experiment is almost identical with that of the distribution of the scores of the entire class. This identity is the effect of chance distribution since other considerations entering into the placing of the students in the classes used in the experiment made selection impossible.

Figures 4 and 5 show the distribution of scores of the county and city groups respectively. The striking difference between the two groups might be of significance in the results of the experiment and should be borne in mind. Why the city group should score so much higher than the rural group is a matter of interest but not of importance to the present study, but the fact that these two groups do differ so widely is of importance in weighing the results of the experiment.

Figures 6 and 7 show the distribution of the two groups of students participating in the experiment during the second and third terms. The groups taken as a whole average approximately the same but the group of the third term is more variable, Q_1 being considerably lower and Q_3 considerably higher than the corresponding points in the second term group.

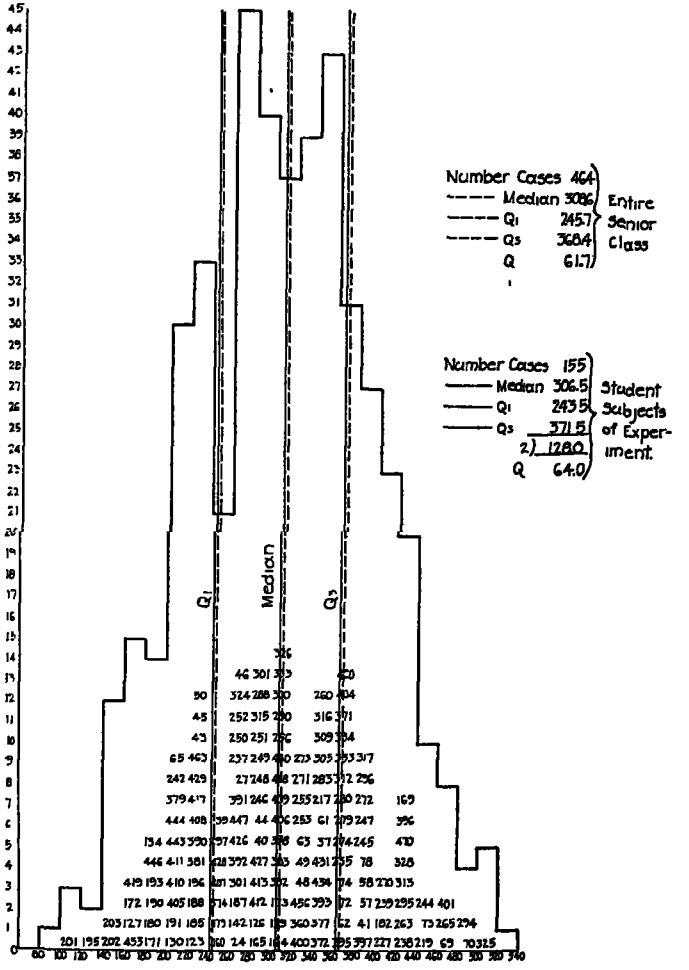

FIGURE 3.—Distribution of I. E. R. Intelligence Score of the 155 student subjects of the experiment compared with the distribution of entire senior class—464 in number.

The numbers appearing in the distributions in Figures 3 to 10 are key designations to particular individuals in the respective groups.

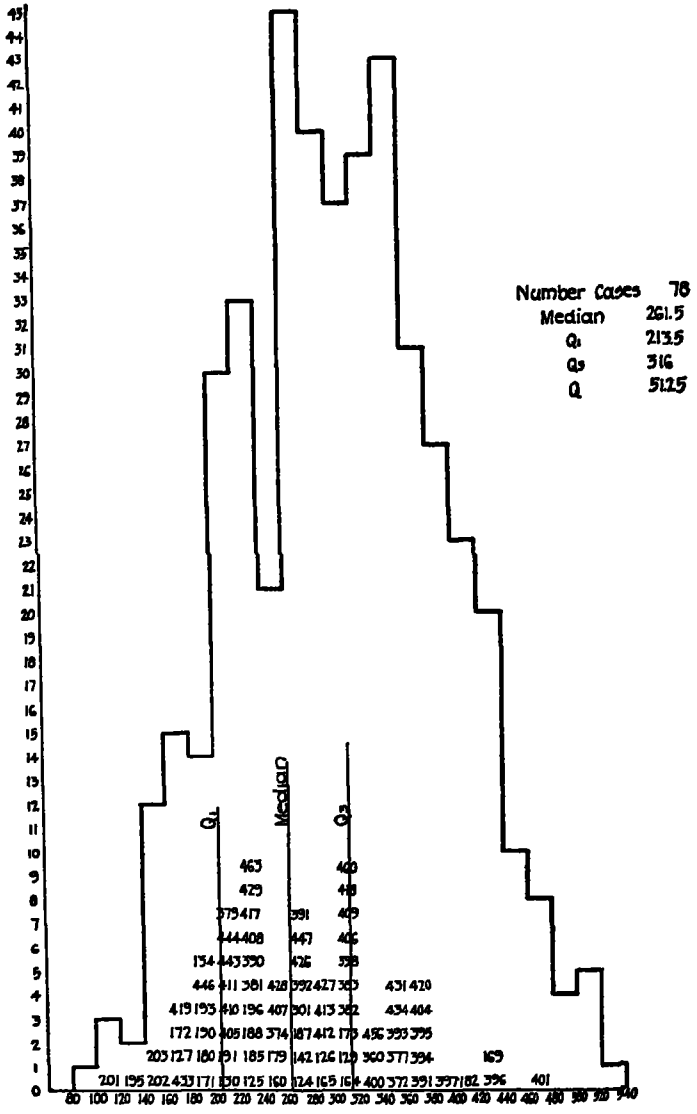

FIGURE 4.—Distribution of I. E. R. Intelligence Scores of the 78 student subjects of the rural group compared with the distribution of entire senior class—464 in number.

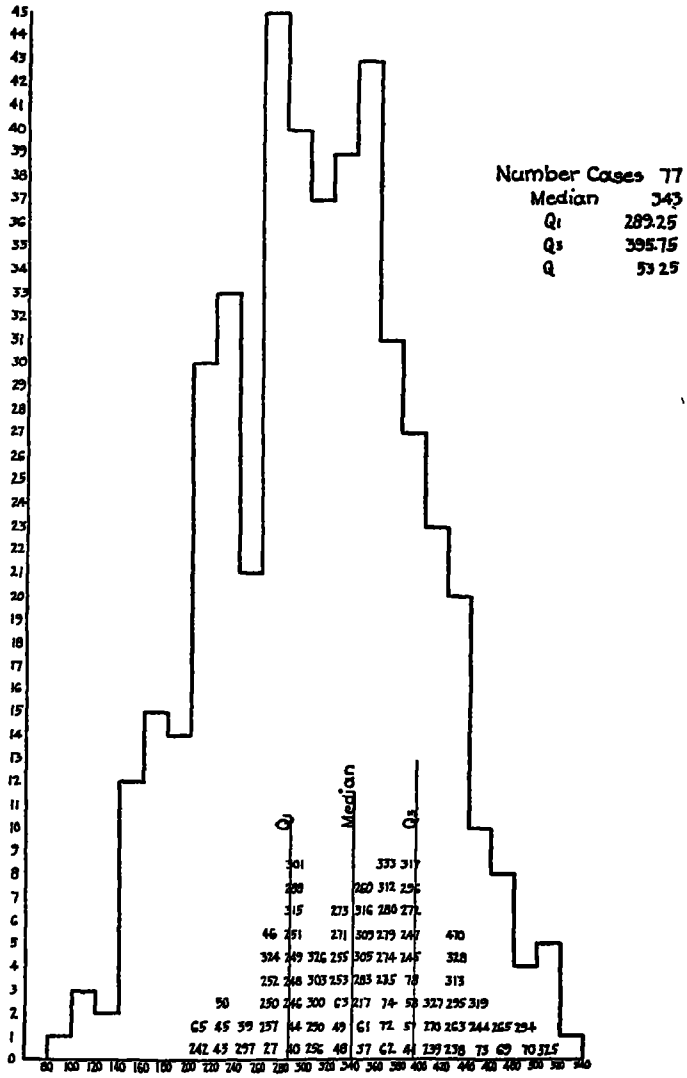

FIGURE 5.—Distribution of I. E. R. Intelligence Scores of the 77 student subjects of the city group compared with distribution of entire senior class—464 in number.

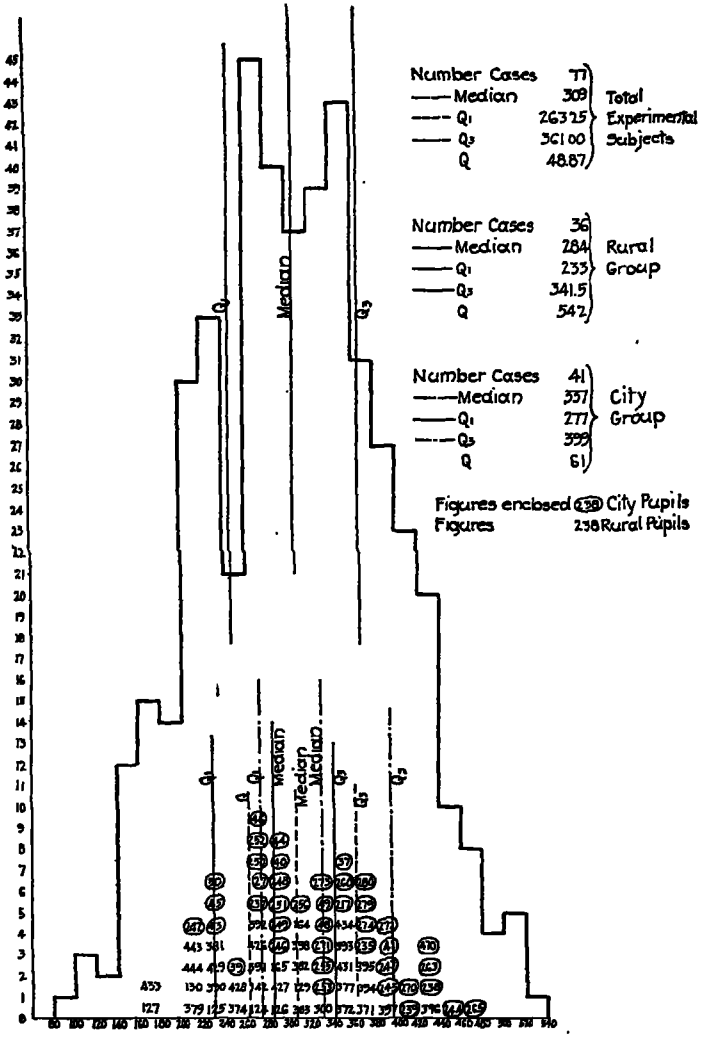

FIGURE 6.—Distribution of I. E. R. Intelligence Scores of 41 city students and of 36 rural student subjects during the second term of student teaching compared with each other and with the distribution of the total class.

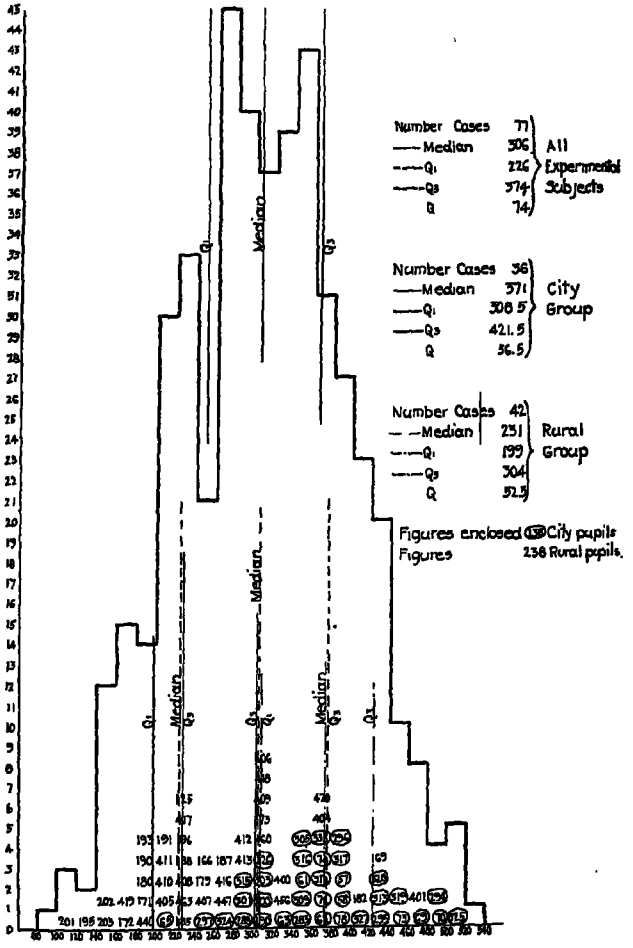

FIGURE 7.—Distribution of I. E. R. Intelligence Scores of 36 city student subjects and of 42 rural student subjects during the third term of student teaching compared with each other and with the distribution of the total class.

The greater variability of the third term group is not due, however, to greater variability within the rural and the city groups taken as units. On the contrary, each of these groups, the rural group particularly so, shows less variability. The difference is due to the fact that the rural group teaching during the third term scored on the average lower and the city group higher than those teaching in the second term. While in each term the entire distribution of the student subjects of the experiment corresponds to the distribution of the entire class, and, therefore, can be considered a typical group, neither the rural nor the city groups as such can be considered typical of the class as a whole, the rural averaging lower and the city higher—particularly in the third term—than the class as a whole.

The age distribution of the student subjects of the experiment as compared with the age distribution of the entire class is shown in Figure 8. It is seen from this figure that the age distribution of the student subjects of the experiment is similar to that of the age distribution for the entire class.

Figures 9 and 10 show the age distributions for the rural and city student subjects of the experiment. The third quartile point is a year higher than that of the city group due mainly to the presence in the rural group of a number of experienced teachers, who after teaching in the state come to the State Normal School to complete the preparation which will entitle them to first-class state certificates. It was thought that the superior scores of the city students in the I. E. R. Intelligence Test might be due to the greater maturity of these students since the elementary and high school course is an 11-year course in the state and a 12-year course in the city. The year less of preparation does not seem, however, to have brought with it a corresponding decrease of a year in average age. Again, as in the case of the intelligence test results, a field for further inquiry is suggested but will not be pursued further here.

It was desired to make comparison of the scholarship during the junior year of the rural and city groups but it was practically impossible to do this with any degree of accuracy.

The experiment was conducted during the year that the merger between the Baltimore City Training School and the Maryland State Normal School was effected. Since, therefore, the records of the junior year of the student subjects of the experiment had been made in different situations and

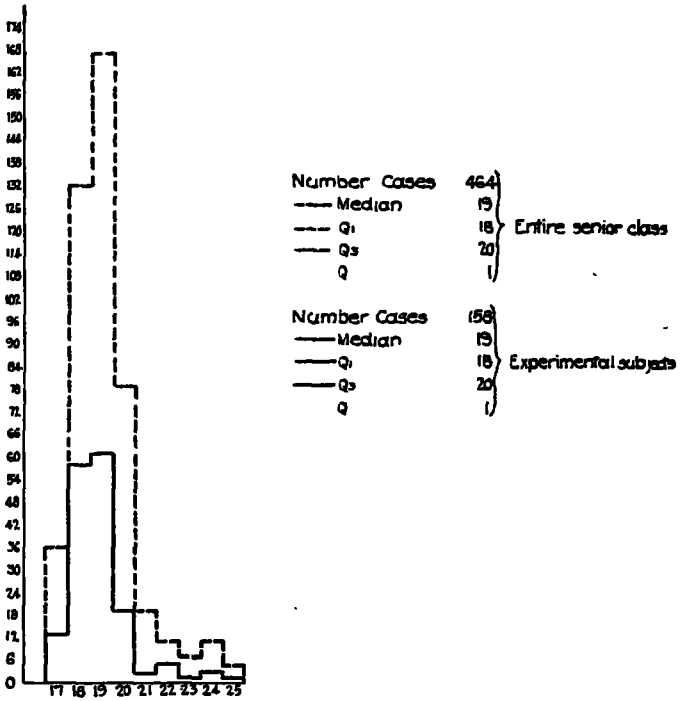

FIGURE 8.—Distribution of the chronological ages of the 158 student subjects of the experiment compared with the distribution of the entire senior class—464 in number.

were made according to entirely different marking systems, the results would not have been comparable.

The status of the student teachers may be summarized from the foregoing as follows:

1. The experiment was conducted during the second and third terms of the year 1924-25 involving 79 student teachers in each term.

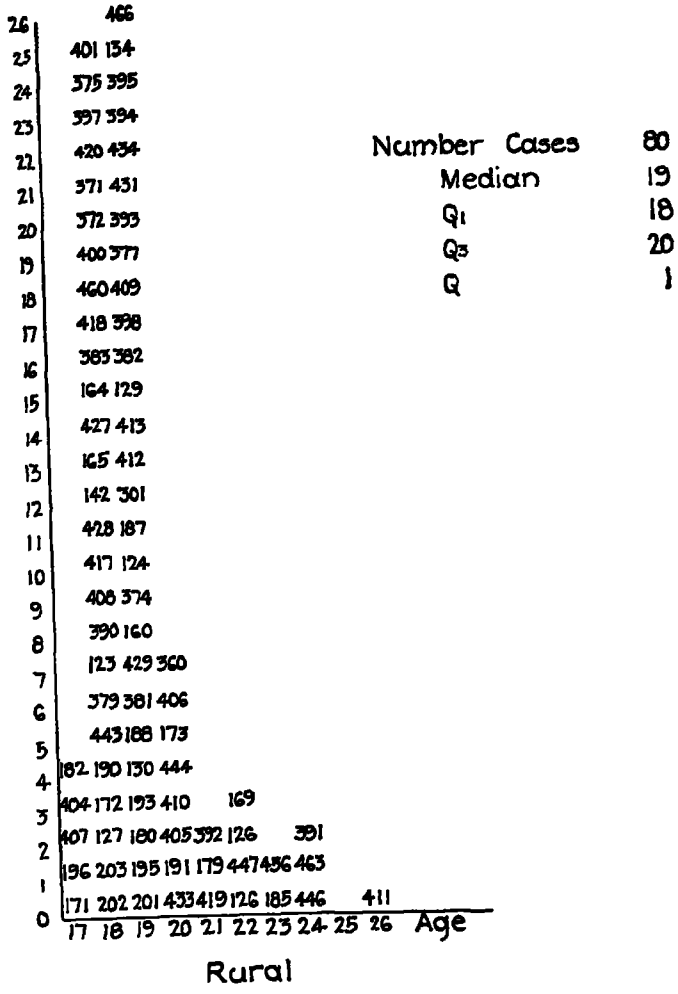

FIGURE 9.—Distribution of chronological ages of 77 student subjects of the rural group.

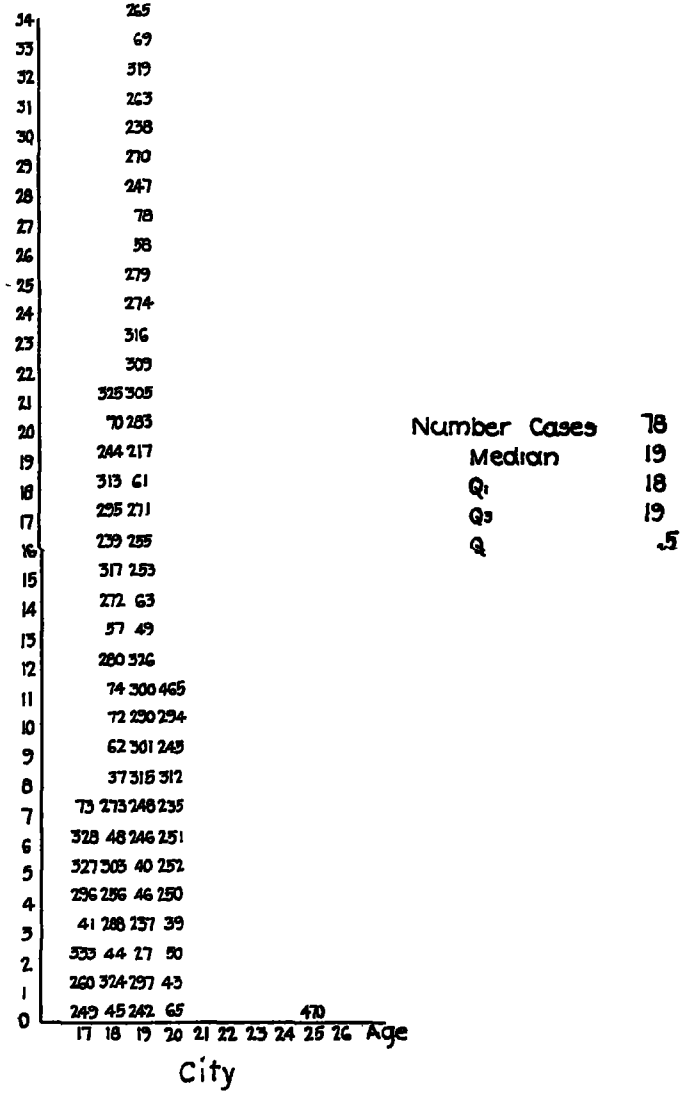

FIGURE 10.—Distribution of chronological ages of 77 student subjects of the city group.

2. The distribution of intelligence scores of the entire group participating in the experiment is typical of the distribution of the entire class.

3. The city group represents more homogeneous training than the rural group.

4. The city students rank higher in intelligence scores than the rural students, the difference being greater in the third term than in the second.

5. The average and the lowest quartile point in the chronological age distribution are the same for the rural and city groups, but the rural group included a few older teachers which made the third quartile point a year higher in that group.

6. The chronological age distribution of the group participating in the experiment is typical of the distribution of the entire class.

CHAPTER IV

Equating the Classes

1. When equated

The classes were equated twice: first, at the beginning, and second, at the end of the experimental teaching period. The groups were equated in the beginning primarily to lessen, as far as possible, the loss of cases that would necessarily be sustained in the final equating. Equating at the beginning had the additional advantage of securing two groups which might be expected, if all conditions in these groups were kept as uniform as possible, to progress at approximately equal rate. The two groups were equated in the end in order to secure two groups of equal initial status in each of the subjects in which experimental teaching was done in order to measure as accurately as possible the relative progress which the experimental and control groups made in these subjects.

2. Equating at the beginning of the experiment

The three groups of schools used, city, rural, and campus, presented three different problems in equating. In the city schools, where two classes of the same grade were in charge of one teacher of practice, these two classes were reorganized into equivalent classes. At The Maryland State Normal Elementary School, where one class only was in charge of a teacher of practice, each of the three classes used was divided into two equivalent groups. In the rural schools, the groups were too small to permit of the division of a single class into two groups; neither was it possible, since the schools were located in different neighborhoods, to secure equality by shifting the children from one school to another as they were shifted from one class to another in the city schools. The best that could be done was to match three of the six rural

schools against the remaining three, aiming to have the two pairs as nearly equal as possible.

In the city and campus schools, each pair of experimental and control groups was under the supervision of the same teacher of practice; in the rural schools, under the supervision of different teachers of practice. This condition introduced a variable into the rural situation which did not exist in the city and campus schools. Since the direct supervision of the experimental teaching was in charge of the teacher of practice, the importance of this variable looms large. When we add to the existence of this variable the condition described above, i. e., that no shifting from school to school to equate the groups could be carried out, it is apparent that conditions in the control and experimental classes in the rural situation were not so nearly equivalent as in the city and campus schools.

Equating in the beginning in the city and campus schools, where it was possible to shift the pupils, was done mainly upon the basis of mental age. If, it was assumed, the main criterion in this experiment is the growth of the pupils along certain specific lines, then the control and experimental groups will be started with most nearly equal chances for success if the groups are equivalent in capacity. It was further assumed that the best measure of capacity available was the mental age of the pupils. When, however, equating upon this basis gave marked discrepancies, first, in chronological age and, second, in educational age, further shifting was done upon these two bases successively. Finally, when shifting upon these bases left a preponderance of one sex or another in either group, further shifting was made to equalize this condition. In all of the shifting done, care was taken to maintain the original balance of the two groups in mental age. In each case, the central tendency as expressed by the arithmetic mean and the variability as expressed by S. D. were considered. The status of the experimental and control groups in mental, educational and chronological ages is shown in Table I.

TABLE I

MENTAL, EDUCATIONAL, AND CHRONOLOGICAL AGES (MEAN AND S. D.) OF EXPERIMENTAL AND CONTROL CLASSES AFTER INITIAL EQUATING

Class	Experimental			Control		
	Cases	Mean	S. D.	Cases	Mean	S. D.
MENTAL AGE						
City Classes						
A	42	92.0	11.6	42	91.7	11.1
B	35	93.8	11.4	35	94.1	9.9
C	37	109.0	14.0	38	109.1	17.4
D	48	114.1	15.1	49	115.3	11.2
E	41	129.1	13.3	44	128.3	14.1
F	33	133.2	15.1	34	132.6	14.7
Campus Elementary Classes						
G	13	111.1	16.5	12	112.4	13.5
H	15	124.3	16.8	14	125.7	15.7
I	14	150.0	18.3	15	149.4	17.9
Rural Classes						
JK	35	113.7	16.2	34	113.1	17.9
L	18	122.8	14.7	12	124.2	19.4
M	33	126.1	18.3	33	128.4	14.5
EDUCATIONAL AGE						
City Classes						
A	37	106.2	9.7	39	105.8	8.7
B	28	107.4	9.7	28	107.2	8.5
C	31	120.4	7.2	34	122.3	6.4
D	45	119.6	8.1	47	119.2	7.2
E	38	132.3	7.7	42	133.0	8.4
F	27	135.0	10.3	25	137.6	11.1
Campus Classes						
G	7	109.5	8.6	9	113.6	9.9
H	13	127.4	16.5	13	125.6	13.5
I	13	137.6	11.4	13	136.7	8.6
Rural Classes						
JK	26	126.5	16.1	22	119.9	13.1
L	15	126.1	16.2	18	126.1	14.1
M	30	132.7	11.0	24	134.5	8.6
CHRONOLOGICAL AGE						
City Classes						
A	44	110.3	12.3	44	109.6	12.2
B	36	102.7	5.5	35	101.7	5.9
C	38	119.2	12.2	39	117.9	12.3
D	49	115.6	9.6	49	115.5	9.5
E	42	135.2	11.6	43	137.6	12.6
F	35	129.5	10.7	36	129.7	11.2

TABLE I.—*Continued*

Class	Experimental			Control		
	Cases	Mean	S. D.	Cases	Mean	S. D.
			Campus Classes			
G	13	108.3	10.4	12	106.0	7.3
H	15	114.0	11.1	14	117.7	13.7
I	14	121.1	17.1	15	124.0	13.4
			Rural Classes			
JK	35	131.0	22.8	36	131.8	21.6
L	18	132.1	16.4	13	127.3	17.7
M	33	134.0	21.7	35	132.8	14.3

3. EQUATING AT THE END OF THE EXPERIMENT

Equating at the end of the experiment was made primarily upon the basis of initial status in the various achievement tests given. Since the main criterion of the experiment is the growth made by the two groups in spelling, arithmetic, and reading, it was of the utmost importance that the initial status of the pupils in the tests measuring these subjects be clearly established. It was difficult to establish this status in the present instance because of the unusually large loss of cases. This was due to several conditions. First, it had been decided, for the sake of greater reliability of results, that the experimenter herself should give all tests. This meant a long period of testing and a consequent shortening of the time allotted to the experiment. The experimenter could not go back to the schools to test absentees without further shortening of the time of the experiment. Attempt was made to have the training teachers do this extra testing, but with the many demands made upon them this was often impossible. Hence, at the beginning many cases were lost. At the end, in order to continue the experiment as long as possible, the testing program was given just before the close of the term. This meant a further loss of cases among children, particularly in the rural districts, who leave school early to work in the summer. Furthermore, the city schools have mid-year promotions when pupils had to be removed from and others had to be brought in to the classes used in this experiment. In the light of this experience, it would have been far better to have had the initial tests given by several testers simultaneously with the

opportunity thus afforded for the testing of absentees and the longer experimental period. The method had to be used in the end. It would have been better to have used it in the beginning also.

With the loss of cases in mind, it was decided that the final equating should be made in two ways:

1. The groups should be equated on the basis of each of the three reading measures, the two spelling measures, the two measures of arithmetic reasoning, and the three measures of arithmetic calculation.

2. The groups should be equated on the basis of composite scores of the tests in reading comprehension, in spelling, in arithmetic reasoning, and in arithmetic calculation.

The first method has the advantage of retaining a comparatively large number of cases, but the results as expressed by any single test cannot be considered of much reliability. The second method provides for a compact, summary measure of growth in each of the four subjects and gives much more reliable results, but it has the disadvantage of retaining a smaller number of cases.

Equating According to Separate Tests.—In every case, extreme scores were eliminated and each score in the experimental group matched approximately with a score in the control group so that the average, range, and standard deviations of each pair of experimental and control classes are almost identical.[1] Table II shows the results when pupils are combined, regardless of the type of school in which the classes were located, into grades. Table III shows the results when pupils are combined, regardless of grade, according to the school in which the classes were located. These two combinations were made, not only for the advantages of the larger number of cases thus obtained, but also that differences in results either in types of schools—campus, city, or rural—or in the grades might be brought out clearly and, in so far as possible, accounted for in the light of conditions existing in one situation or another.

[1] The original data for the classes taken individually are on file in the library of The Johns Hopkins University.

TABLE II

EXPERIMENTAL AND CONTROL GROUPS EQUATED BY GRADES—3, 4, AND 5—INITIAL STATUS IN READING, ARITHMETIC AND SPELLING

Experimental				Control				
Cases	Range	Av.	S. D.	Cases	Range	Av.	S. D.	Test
				GRADE 3				
				Reading				
66	78–133	105.30	12.55	66	78–133	105.30	12.55	Stanford
70	22–59	6.80	6.80	70	22–57	36.15	7.70	Thorndike / McCall
56	0–9	2.48	2.48	56	0–9	5.43	2.50	Monroe
				Arithmetic Calculation				
69	1.6–15	9.20	3.72	69	2.3–21.3	9.9	3.87	Stanford
62	20–64	44.40	10.15	62	20–64	44.8	9.40	Monroe
50	0–39	20.05	8.05	50	6–31	19.05	7.15	Courtis
				Arithmetic Problems				
62	0–44	29.35	11.60	62	0–45	30.70	12.00	Buckingham
60	0–48	26.75	10.40	60	0–48	27.10	10.30	Stanford
				Spelling				
72	2–25	12.81	4.11	72	2–23	13.14	4.14	Morrison
57	6–82	46.15	14.50	57	6–78	46.15	14.00	Stanford
				GRADE 4				
				Reading				
80	97–157	123.00	11.75	80	98–157	123.25	10.00	Stanford
89	29–55	39.65	5.55	89	26–55	39.90	5.40	Thorndike
72	3–12	8.23	2.72	72	3–13	8.12	2.76	Monroe
				Arithmetic Calculation				
80	7–44.6	22.2	9.36	80	6.3–45.3	21.72	9.51	Stanford
88	36–92	70.55	12.50	88	36–92	69.35	12.90	Monroe
86	6–92	51.25	21.80	86	8–81	51.48	20.95	Courtis
				Arithmetic Problems				
83	0–57	39.30	6.80	83	0–51	39.05	7.30	Buckingham
87	8–72	39.80	12.50	87	8–80	40.40	12.45	Stanford
				Spelling				
86	7–38	19.59	6.93	86	4–36	19.35	6.81	Morrison
83	22–104	63.70	17.20	83	26–120	64.40	18.35	Stanford
				GRADE 5				
				Reading				
61	107–172	139.80	13.25	61	109–173	138.95	12.35	Stanford
84	27–59	44.10	6.40	84	26–57	44.50	5.45	Thorndike
64	4–13	9.35	2.20	64	4–13	9.43	2.13	Monroe

TABLE II.—*Continued*

Experimental				Control				
Cases	Range	Av.	S. D.	Cases	Range	Av.	S. D.	Test
			Arithmetic Calculation					
81	17.1–67.5	38.07	12.09	81	13.3–69.5	37.32	11.49	Stanford
79	52–116	86.50	12.20	79	48–124	87.50	13.35	Monroe
71	50–190	91.60	26.40	71	43–190	93.75	25.20	Courtis
			Arithmetic Problems					
79	0–63	49.70	8.85	79	0–61	48.85	8.80	Buckingham
75	24–80	55.20	11.60	75	24–88	55.70	12.20	Stanford
			Spelling					
83	15–40	27.39	6.51	83	13–42	27.78	6.27	Morrison
69	60–134	88.20	18.70	69	50–138	87.45	20.15	Stanford

TABLE III

GROUPS EQUATED AT CLOSE OF EXPERIMENT, AFTER CLASSES HAVE BEEN COMBINED ACCORDING TO TYPE SCHOOL—CITY, CAMPUS, OR RURAL—GRADES DISREGARDED, ON INITIAL STATUS IN EACH ACHIEVEMENT TEST, AND AFTER ELIMINATION HAD BEEN MADE BECAUSE OF ABSENCE AT ONE OR ANOTHER OF THE TESTS GIVEN AND BECAUSE OF EXTREME SCORES

Experimental				Control				
Cases	Mean	Range	S. D.	Cases	Mean	Range	S. D.	Test
			CITY CLASSES					
			Reading					
158	119.90	78–172	18.80	158	120.05	72–173	18.00	Stanford
170	38.65	22–59	7.30	170	38.60	22–57	6.90	Thorndike
142	7.09	0–13	3.02	142	7.08	0–13	3.07	Monroe
			Spelling					
152	65.15	6–134	24.45	152	64.45	6–138	25.60	Stanford
177	19.11	2–40	8.13	177	19.08	2–42	7.89	Morrison
			Arithmetic Calculation					
163	66.25	20–100	19.20	163	66.35	20–105	19.30	Stanford
164	22.41	0–59	14.22	164	22.44	0–61	13.62	Monroe
133	55.34	7–149	16.13	133	55.63	6–147	14.95	Courtis
			Arithmetic Problems					
154	39.30	4–80	16.20	154	39.60	4–88	16.35	Buckingham
164	36.85	0–59	12.05	164	36.53	0–61	11.75	Stanford
			CAMPUS CLASSES					
			Reading					
27	130.45	96–159	18.05	27	128.25	97–159	16.90	Stanford
31	44.10	31–55	6.85	31	44.60	28–55	7.00	Thorndike
19	8.58	1–13	3.15	19	8.37	1–13	3.05	Monroe

TABLE III.—*Continued*

Experimental				Control				
Cases	Mean	Range	S. D.	Cases	Mean	Range	S. D.	Test
			Spelling					
28	70.20	30–116	24.75	28	70.95	32–120	20.75	Stanford
31	22.02	7–36	9.06	31	22.32	6–40	8.88	Morrison
			Arithmetic Calculation					
29	74.45	30–100	22.05	29	73.00	20–100	22.15	Stanford
29	24.99	3–54	13.29	29	23.52	3–51	12.21	Monroe
33	46.96	5–112	15.34	33	47.60	10–98	13.62	Courtis
			Arithmetic Problems					
29	44.00	20–72	16.35	29	45.85	20–72	13.25	Buckingham
30	43.80	0–57	10.30	30	44.00	0–57	11.10	Stanford
			RURAL CLASSES					
			Reading					
22	125.50	109–146	9.15	22	129.20	112–158	8.80	Stanford
41	40.90	28–55	5.90	41	41.30	28–55	6.60	Thorndike
31	7.48	3–11	2.30	31	7.32	3–12	2.34	Monroe
			Spelling					
36	79.60	44–112	20.90	36	79.70	44–110	19.15	Stanford
33	25.50	9–36	7.29	33	25.68	9–34	6.51	Morrison
			Arithmetic Calculation					
38	76.70	20–115	19.60	38	77.10	20–120	22.40	Stanford
39	30.96	0–63	16.58	39	31.95	0–69	17.76	Monroe
41	74.80	0–190	30.48	41	71.53	10–190	26.53	Courtis
			Arithmetic Problems					
39	46.35	20–80	14.15	39	48.00	16–84	15.30	Buckingham
32	43.00	0–63	10.95	32	43.00	0–61	10.95	Stanford
			TOTAL CLASSES					
			Reading					
207	122.40	78–172	18.40	207	122.10	72–173	17.20	Stanford
242	39.70	22–59	7.30	242	39.90	22–57	7.40	Thorndike
192	7.31	0–13	2.97	192	7.25	0–13	2.95	Monroe
			Spelling					
216	68.15	6–138	23.75	216	68.05	6–134	24.50	Stanford
241	20.13	2–40	8.25	241	20.37	2–42	8.37	Morrison
			Arithmetic Calculation					
230	69.00	20–115	20.40	230	68.95	20–120	20.65	Stanford
232	23.52	0–63	14.76	232	23.88	0–69	14.43	Monroe
207	57.86	0–190	19.87	207	57.06	6–190	17.75	Courtis
			Arithmetic Problems					
222	41.50	4–80	16.20	222	41.85	4–88	16.40	Buckingham
226	40.45	0–63	11.90	226	40.20	0–61	11.70	Stanford

Because of the large loss of cases sustained in compiling composite scores, it seemed futile to attempt the equation of individual classes. Instead, the classes were combined according to grade and then equated. Two main considerations entered into the problem of equating by grades:

1. The method to be used in transmuting the scores of the various tests in each of the four subjects into comparable measures.

2. The amount of weight to be given to the various transmuted measures.

1. Comparable Measures: Percentile Method.—The tests in each subject—reading, spelling, arithmetic calculation, and arithmetic problems—were combined to form a composite score. In each case, the scores were transmuted into age scores using the Stanford age for each subject as the basis for transmutation. The method used in reading in Grade 3 will be described as an example:

The Stanford reading ages for Grade 3 were plotted into a percentile curve. Likewise the Monroe Reading Scores and the Thorndike McCall T scores were plotted. The pairs of corresponding scores for the various percentile ranks for Standford and Thorndike McCall and for Stanford and Monroe, were then plotted and lines of relationship drawn. From the lines of relationship, tables of correspondence for Stanford and Thorndike McCall and for Stanford and Monroe were then made. It was possible to determine from these tables the scores in Monroe and Thorndike McCall equivalent to the Stanford Reading Age. The sum of the three scores then formed the unweighted composite reading score for each pupil.[1]

2. Weighting.—The opinion of four experts in the field of educational measurements was sought in making the decision as to the weight to assign to the various tests employed. In general, the opinion was in favor of giving more weight to the Stanford Achievement Test than to the other tests. In

[1] Otis, Arthur S. Statistical Method in Educational Measurement. pp. 101-117.

the case of each test, an approximate average of the four weights was found and assigned as follows:

READING

	Stanford	Thorndike McCall	Monroe
	4	3	1.2
	3	1	1
	4	2	1
	4	2	1
Approximate Average	4	2	1

SPELLING

	Stanford	Morrison McCall
	1	1
	1	1
	2	1
	2	1
Approximate Average	2	1

ARITHMETIC CALCULATION

	Stanford	Monroe	Courtis
	1	1	
	2	1	
	2	1	1
	2	1	1
Approximate Average	2	1	1

ARITHMETIC PROBLEMS

	Stanford	Buckingham
	2	3
	2	1
	2	1
	2	1
Approximate Average	2	1

The standing of the equated classes grouped by grades in terms of mean and S. D. is shown for the unweighted composite scores in Table IV. Table V gives the standing in the weighted scores when the classes are grouped according to grade.

Attempt was made to equate according to type schools—city, campus, and rural—as was done in the case of the individual tests (Table III). The large variability found when

TABLE IV

GRADES EQUATED AT CLOSE OF EXPERIMENT ON BASIS OF INITIAL STATUS IN COMPOSITE SCORES IN READING, SPELLING, ARITHMETIC CALCULATION, AND ARITHMETIC PROBLEMS, UNWEIGHTED

	Experimental				Control			
	Cases	Mean	Range	S. D.	Cases	Mean	Range	S. D.
READING								
Grade 3	29	314.86	267–371	28.45	29	313.56	264–368	26.65
Grade 4	39	359.36	308–417	23.46	39	360.59	307–417	20.02
Grade 5	33	415.60	347–469	26.50	33	413.12	349–470	23.82
SPELLING								
Grade 3	35	89.05	53–138	19.46	35	88.71	53–136	19.58
Grade 4	51	123.30	72–206	33.00	51	122.86	61–214	33.48
Grade 5	43	165.48	118–250	32.60	43	165.52	117–253	32.33
ARITHMETIC CALCULATION								
Grade 3	27	137.33	109–169	12.39	27	138.66	108–170	15.39
Grade 4	40	208.07	158–258	24.45	40	205.28	154–252	23.77
Grade 5	37	263.62	214–315	16.40	37	262.00	215–306	15.67
ARITHMETIC PROBLEMS								
Grade 3	33	51.70	20–84	16.40	33	51.00	23–80	15.67
Grade 4	57	79.77	28–146	21.31	57	79.62	34–154	20.70
Grade 5	55	119.38	82–170	15.64	55	119.25	86–176	14.51

TABLE V

COMPARISON OF EXPERIMENTAL AND CONTROL SITUATIONS—GRADES DISREGARDED—IN INITIAL STATUS IN COMPOSITE SCORES, UNWEIGHTED

	Experimental				Control			
	No.	Mean	Range	S. D.	No.	Mean	Range	S. D.
READING								
City	85	361.31	267–452	46.81	81	360.95	264–440	44.84
M. S. N.	11	382.10	305–469	54.75	12	388.25	318–470	53.88
Rural	6	356.84	308–395	31.41	6	360.25	317–386	20.15

TABLE V.—*Continued*

	Experimental				Control			
	No.	Mean	Range	S. D.	No.	Mean.	Range	S. D.
				SPELLING				
City	98	126.83	53 265	44.75	93	122.68	53 261	42.76
M. S. N.	19	160.00	59 309	75.32	18	175.61	89 255	45.35
Rural	12	238.33	124 314	55.88	18	200.56	124 276	39.03
				ARITHMETIC CALCULATION				
City	76	203.48	109 304	53.56	76	198.77	108 302	54.14
M. S. N.	19	240.58	120 350	62.07	16	233.45	132 287	38.46
Rural	9	277.78	158 327	46.14	12	236.92	145 306	43.01
				ARITHMETIC PROBLEMS				
City	107	79.82	20 147	30.62	109	84.36	23 176	31.90
M. S. N.	22	110.91	35 191	39.49	18	105.78	40 139	26.62
Rural	16	136.38	56 218	36.16	18	93.95	34 154	29.42

several grades were considered in one grouping made equating impossible without reducing the already small number of cases to a point where no significance whatever could be attached to the results. On the other hand, because of the differences in the conditions under which the experiment was conducted, it seemed of importance to get some composite measure of the status of the three type groups in order that the results might be studied from the angle of the differences in the conditions of these groups. Accordingly, the status in the composite scores, unweighted and weighted, are presented in Tables V and VII respectively. No attempt at equating the groups was made, the scores of the pupils selected for equating by grade merely being redistributed according to type schools.

While in the final equating initial status in the subjects whose gains in the experimental and control groups were to be compared was taken as the main basis for equating, mental age and chronological age were, as in the initial equating, also considered. The results are shown in Tables VIII and IX.

TABLE VI

GRADES EQUATED AT CLOSE OF EXPERIMENT ON BASIS OF INITIAL STATUS IN COMPOSITE SCORES IN READING, SPELLING, ARITHMETIC CALCULATION AND ARITHMETIC PROBLEMS, WEIGHTED

	Experimental				Control			
	Cases	Mean	Range	S. D.	Cases	Mean	Range	S. D.
READING: WEIGHT: STANFORD—4; MONROE COMPREHENSION—1; THORNDIKE MCCALL—2								
Grade 3	29	737.93	591 873	67.02	29	738.93	613 862	64.47
Grade 4	39	835.82	712 1016	67.23	39	843.47	744 960	45.81
Grade 5	33	977.12	830 1112	63.47	39	968.61	826 1108	57.97
SPELLING: WEIGHT: STANFORD—2; MORRISON MCCALL—1								
Grade 3	35	133.66	77 220	33.51	35	135.85	89 208	29.64
Grade 4	51	185.25	108 310	50.47	51	185.49	87 334	51.20
Grade 5	43	259.51	180 399	52.55	43	262.28	178 387	53.79
ARITHMETIC CALCULATION: WEIGHT: STANFORD—2; COURTIS—1; MONROE—1								
Grade 3	27	184.63	145 229	19.21	27	186.97	144 226	19.13
Grade 4	40	278.25	217 350	32.64	40	271.55	194 336	37.22
Grade 5	37	350.14	275 415	38.02	37	348.89	268 402	35.75
ARITHMETIC PROBLEMS: WEIGHT: STANFORD—2; BUCKINGHAM—1								
Grade 3	33	77.66	24 124	25.82	33	77.88	35 124	24.02
Grade 4	57	119.10	55 218	32.60	57	119.94	56 234	32.56
Grade 5	55	174.91	114 250	26.16	55	176.49	114 264	24.99

TABLE VII

COMPARISON OF EXPERIMENTAL AND CONTROL SITUATIONS—GRADES DISREGARDED IN INITIAL STATUS IN COMPOSITE SCORES, WEIGHTED

	Experimental				Control			
	No.	Mean	Range	S.D.	No.	Mean	Range	S.D.
READING: WEIGHT: STANFORD—4; THORNDIKE MCCALL—2; MONROE—I								
City	85	845.67	591 1071	113.38	81	846.16	613 1040	104.46
M. S. N.	10	920.60	773 1112	122.27	12	910.17	706 1108	127.44
Rural	6	844.67	736 934	69.58	8	849.13	759 904	42.98
SPELLING: WEIGHT: STANFORD—2; MORRISON MCCALL—I								
City	98	180.73	77 399	68.54	93	185.70	87 387	63.93
M. S. N.	19	197.05	91 309	68.58	19	222.42	89 375	69.63
Rural	12	230.34	59 314	71.83	17	217.12	85 384	82.38
ARITHMETIC CALCULATION: WEIGHT: STANFORD—2; COURTIS—I; MONROE—I								
City	76	271.18	145 408	69.03	76	264.68	144 402	71.12
M. S. N.	19	294.58	152 408	78.15	16	314.81	188 387	51.69
Rural	19	317.33	214 415	61.37	12	311.08	193 402	56.95
ARITHMETIC PROBLEMS: WEIGHT: STANFORD—2; BUCKINGHAM—I								
City	107	104.36	24 161	27.68	109	126.21	35 264	47.62
M. S. N.	22	116.73	59 191	31.57	18	153.23	64 211	40.80
Rural	16	160.62	56 250	48.52	18	143.61	58 234	44.40

TABLE VIII

MENTAL AGE—FINAL EQUATING OF EXPERIMENTAL AND CONTROL GROUP BY GRADES

	Experimental			Control		
	Cases	Mean	S. D.	Cases	Mean	S. D.
			GRADE 3			
Reading	29	94.22	10.75	29	96.60	10.45
Arith. Cal.	27	94.53	13.20	27	96.57	8.85
Arith. Prob.	33	94.01	12.65	33	95.23	10.85
Spelling	35	91.64*	12.90	35	95.64*	14.25
			GRADE 4			
Reading	39	120.58*	11.20	39	111.99*	11.50
Arith. Cal.	40	117.05	15.60	40	117.52	14.00
Arith. Prob.	57	117.52	15.85	57	117.47	13.80
Spelling	51	117.31	15.75	51	117.50	13.20
			GRADE 5			
Reading	33	133.53	13.90	33	134.46	12.90
Arith. Cal.	37	133.85	17.30	37	136.82	14.60
Arith. Prob.	35	133.50	16.40	35	137.32	15.25
Spelling	43	134.70	16.80	43	136.57	14.40

* Instances in which Equating is not close; See p. 52

TABLE IX

CHRONOLOGICAL AGE—FINAL EQUATING OF EXPERIMENTAL AND CONTROL GROUP BY GRADES

	Experimental			Control		
	Cases	Mean	S. D.	Cases	Mean	S. D.
			GRADE 3			
Reading	29	113.66	12.45	29	113.36	12.40
Arith. Cal.	27	102.74	7.05	27	101.98	7.05
Arith. Prob.	33	113.83*	9.75	33	108.04*	8.40
Spelling	35	107.528	9.60	35	107.81	7.30
			GRADE 4			
Reading	39	123.66	10.05	39	123.60	10.25
Arith. Cal.	40	111.15*	11.40	40	117.10*	11.65
Arith. Prob.	57	111.728*	22.25	57	117.413*	9.85
Spelling	51	117.54	11.15	51	117.324	11.25
			GRADE 5			
Reading	33	127.05	13.85	33	126.71	12.50
Arith. Cal.	37	127.09	15.15	37	125.96	13.95
Arith. Prob.	35	127.12	14.00	35	126.78	12.50
Spelling	43	126.71	13.70	43	126.34	14.55

* Instances in which Equating is not close; See p. 52

In equating according to chronological age (Table IX), it will be noted that there are 3 cases in which the equating is not close (Note, Table IX). Since one of these favors the experimental and two favor the control, it was thought best to leave the discrepancy rather than destroy the equating according to initial status.

Again, in equating for mental age there are two discrepancies (Note, Table VIII). Since, however, one of these again favors the experimental and the other favors the control group, it was thought best not to disturb the equating according to initial status. The more serious discrepancy is in the case of reading, Grade 4. In attempting to make a better equating, it was found that the discrepancy was due to one high mental age in the experimental and to one low mental age in the control. Removing these cases resulted in much better equating according to mental age:

Experimental:	112.051	Mean;	12.00 S. D.
Control:	111.85	Mean;	10.90 S. D.

If this were done, however, the equating according to initial status would not be so good, resulting in:

Experimental:	831.81	Mean;	59.20 S. D.
Control:	850.84	Mean;	45.00 S. D.

in place of the original equating:

Experimental:	835.582	Mean;	67.23 S. D.
Control:	843.47	Mean;	45.81 S. D.

It was, therefore, decided not to attempt a closer equating according to mental age.

CHAPTER V

Equating Experimental and Control Groups of Student Teachers

I. EQUATING AT THE BEGINNING OF THE EXPERIMENT

Attempt was made to equate the groups of student teachers at the beginning of the experiment. Ordinarily, assignment to training centers at the Maryland State Normal School at Towson is based primarily upon the student's own choice of grade. Effort is also made to apportion strong and weak students in such a way that no training center will suffer the disadvantage of being overburdened with an undue number of the weaker students. Again, the personality of the student in relation to the personality of the teacher of practice is considered in order that students may be placed to the best advantage with respect to their individuality. Every effort was made to assign the students in such way as to equate the experimental and control groups and at the same time carry out the above policy.

As will be shown in the tables, it was possible to carry out the plan with fair success at the first assignment of the students at the beginning of each of the two terms. It is the custom, however, to shift the students at the end of the first half of the term in order to give them two teaching experiences. Here it was found to be impossible to preserve the original composition of the *control* group without serious injustice to the students. This resulted in the loss of certain of the students in the control group by transfers to training centers not included in the experiment and the substitution of students not in the original control groups. The original composition of the experimental groups was preserved even at a sacrifice, in some cases, of the best adjustment for the student. In the third term, for the sake of the experiment, no mid-term shift was made in the city experimental group.

Every experimental student who persisted to the end of the experiment had a complete term of training in the method used in this experiment.

The bases for equating the students were: Intelligence, Scholarship, Prognosis of Teaching Ability, and Age. Intelligence was measured by the I. E. R. Intelligence Test (Chapter III). Scholarship was measured by an average of all the grades of the junior year. A difficulty was encountered here in the different record systems used for the city and the county students, the city students having been marked during their junior year at the Baltimore City Training School on the 0-100 basis, while the county students had been marked on the 1-7 basis. This rendered an accurate comparison of the scholarship of the two groups of city and county students unsatisfactory but served the purpose of equating city experimental and control groups and county experimental and control groups. Attempt was made to secure a purely subjective rating of prognosis of teaching success by three instructors who knew the students quite well. It was possible to secure this at the beginning of the two terms for the city students, but it was possible to secure it only at the beginning of the second term for students assigned to county training centers.

Equating was carried out by matching as nearly as possible the students in an experimental center with its paired control center. In addition, effort was made to equalize the entire group of city experimental students with the entire group of city control students, the county experimental with the county control, and the campus experimental with the campus control.[1]

2. EQUATING AT THE END OF THE EXPERIMENT

The students were equated again at the end of the experiment in order to secure groups of more nearly equal initial status than it was possible to secure at the beginning of the experiment because of the other conditions entering into the assignment of the students indicated above. It was desired

[1] Tables indicating the above equating of groups of student teachers are on file in the library of The Johns Hopkins University.

to equate, also, on the basis of possible teaching ability as indicated by the teaching done by the student teachers during the first week of their student teaching period.

In the summer previous to the beginning of this study a preliminary experiment had been performed for the purpose of selecting a good rating scheme to use in this phase of the work. In brief, the experiment consisted in the evaluation of the work of three demonstration teachers in each of three summer schools by groups of eight supervisors, superintendents and teachers. Each evaluator rated each teacher five times, the first rating being a purely subjective one and the other ratings made with the help of five representative rating cards. The results showed that no greater uniformity of rating of the same teachers by different evaluators was secured with the aid of rating cards than was secured on a purely subjective rating. No doubt, as Johnson has shown, greater uniformity of rating can be secured with the use of a rating card after a period of training of the evaluators.[1] Since such training was impossible in the present experiment and since the limitations of time did not permit the rating of 79 students by a sufficiently small group of evaluators to make such rating of significance, equating by initial rating was omitted.

The students, then, were equated at the end of the experiment first, on the basis of the I. E. R. Intelligence Test, then according to scholarship, and finally according to chronological age, in each case extremes being eliminated. Since in the middle of each term, the campus students and the rural students are interchanged, these two groups were combined in equating. Tables X and XI show the equated experimental and control groups for city and combined campus and rural schools, respectively.

[1] Johnson, F. W. "Supervision of Instruction," *School Review*, XXX, 1922, pp. 742-754.

TABLE X

FINAL EQUATING, STUDENT TEACHERS IN CITY CLASSES

Experimental				Control			
Student key no.	I. E. R. intel.	Junior grades	Chron. age mos.	Student key no.	I. E. R. intel.	Junior grades	Chron. age mos.
45	226	84	218	319	446	87	216
256	311	85	216	46	262	80	229
309	341	86	228	252	272	80	242
288	299	86	219	242	200	83	234
61	340	85	233	40	284	82	217
239	417	84	221	27	271	86	233
297	258	87	231	326	307	84	225
272	397	86	224	62	272	88	221
283	355	83	217	301	287	89	247
58	398	88	237	37	340	90	208
246	288	87	194	274	377	85	227
312	362	88	238	260	343	86	212
63	335	86	231	49	338	80	217
39	247	85	239	50	225	84	246
251	281	80	248	279	367	86	227
69	464	85	235	300	300	88	235
296	392	88	207	316	351	84	227
324	260	84	225	237	266	87	208
248	293	82	237	271	337	88	229
317	388	83	224	255	338	84	239
65	212	81	230	250	273	87	248
328	424	87	210	78	383	85	235
249	287	74	236	303	310	85	204
238	432	88	228	48	334	81	219
72	373	85	219	265	465	89	226
327	419	85	210	280	373	87	204
333	375	87	207	235	363	88	241
43	221	87	244	294	499	88	243
325	515	94	217	313	425	81	225
273	337	88	267	74	370	86	226
41	399	85	248	57	381	88	222
Sum	10646	2656	7036		10459	2646	7258
Av.	343.41	85.67	226.96		337.38	85.35	226.81
S. D.	73.93	2.29	14.11		66.31	2.84	18.05
Num.	31	31	31		31	31	31

TABLE XI

FINAL EQUATING, STUDENT TEACHERS IN CAMPUS AND RURAL CLASSES

Experimental				Control			
Student key no.	I. E. R. intel.	Junior grades	Chron. age mos.	Student key no.	I. E. R. intel.	Junior grades	Chron. age mos.
456	333	*	276	129	316	4.8	211
125	223	4.5	218	397	393	5.6	211
188	226	4.1	225	426	266	3.7	214
429	227	4.3	216	431	341	4.5	224
165	285	4.6	208	398	311	4.5	222
447	272	*	271	407	248	4.7	195
433	173	3.7	226	408	231	3.8	209
164	304	4.4	211	427	291	4.1	207
374	257	3.4	218	416	278	4.5	204
460	304	*	218	142	275	5.3	207
360	322	4.3	229	134	198	3.2	224
180	199	4.0	214	171	196	4.2	199
179	240	4.3	243	383	316	5.2	213
394	371	6.1	216	381	221	4.8	214
443	215	4.6	214	172	167	4.8	222
173	300	5.4	227	196	225	4.9	195
185	236	*	284	202	156	3.9	208
382	307	4.6	226	395	373	5.3	204
126	283	4.2	283	372	344	4.5	207
377	342	6.0	223	130	200	4.0	238
Sum	5419	72.5	4646		5346	91.1	4228
Av.	270.95	4.5	232.30		267.30	4.5	211.4
S. D.	51.35	.73	24.37		66.46	1.96	10.33
Num.	20	16	20		20	20	20

* No junior grades, because these students entered the Maryland State Normal School in their senior year.

CHAPTER VI

The Materials of the Experiment

I. GENERAL PLAN

The experiment was conducted during the second and third terms of the school year 1924-25. During the second term, nine weeks dating from January 5th to March 14th were used. During the third term, ten weeks dating from March 16th to May 23rd were used. While some procedures in the third term were different from some in the second term, the administration in both terms was essentially the same.

1. The experimenter met each of the three groups of students assigned to do the experimental teaching in the city, campus, and rural centers for the purpose of explaining the main features of the experiment. They were told that they would be given the results of standard tests previously given to the children in order that they might base their instruction in arithmetic calculation and problem solution, spelling, and reading upon such results; that specific instructions would be sent them weekly in bulletin form; that they would keep diaries of their work in these subjects upon blanks sent to them; that their teachers of practice would guide them in this phase of their work as in other phases. No mention was made to them of the existence of a control group. On the contrary, every effort was made to prevent such knowledge from coming to them.

2. Nine mimeographed bulletins in the second term and ten such bulletins in the third term were sent to each student, each bulletin indicating one week's work to be carried out.

3. Each week the students sent to the experimenter diaries of their work and any other material indicative of the children's activities as called for by the bulletin.

4. The experimenter made attempt to pay at least one visit to each student, but this was found to be impracticable.

5. More or less informal conferences were held between the experimenter and the teachers of practice on the progress of the work.

The details will be treated separately for the two terms.

2. EXPERIMENT DURING THE SECOND TERM

According to the original plan as indicated in Chapter II, the time devoted to the experimental teaching was to be allotted as follows:

3 30-minute periods per week in reading
3 15-minute periods per week in spelling
3 30-minute periods per week in arithmetic, the time to be divided equally between arithmetic calculation and problem solution

Because of the work involved in scoring the tests and tabulating and diagnosing the results, it was found impossible to launch the program in its entirety in the beginning. The changes made in the original plan will be indicated later. Copies of the instructions for keeping the diary of test-determined instruction, of the diary blank, of a sample score sheet sent to the experimental classes with Bulletin I, and of a sample of the suggestions for grouping the children in reading are found in the appendix (see C, D, E, and F).[1] Summaries of the nine bulletins sent during this term follow:

Bulletin 1 dealt with reading. It set the problems for the week, indicated the amount of time to be used, assigned definite phases of the work to the students, and gave suggestions of type material to be used. In addition, a score sheet of the initial test in reading, together with directions for grouping the children, was sent to each center (Appendix E). The score sheet gave each child's score in the six reading measures available, along with class norms for the preceding grade as well as for the grade which the children had reached at the time of the experiment. The particular types of work outlined in the bulletin suitable to the various groups of children were indicated.

Bulletin 2 contained directions for starting the work in spelling. In addition to scoring the Morrison McCall Spelling

[1] Other samples of score sheets sent with the various bulletins are on file in the library of The Johns Hopkins University.

Test, given among the initial tests at the beginning of the school year, analysis of errors had been made on diagnostic score sheets. These diagnostic score sheets formed the basis for the selection of the words to be used in the spelling exercises given in the bulletin. With the bulletin, the scores made by each child in the Morrison McCall and the Stanford Spelling Tests and also a summary of the number of times each word of the Morrison McCall Test had been misspelled were sent to each experimental class.

Bulletin 3 gave directions for the work in arithmetic calculation. The exercises for the week were based upon the results shown in the Courtis Arithmetic Test and were confined to addition. With this bulletin, scores for each child in the various tests given and also the detailed score sheet of the Courtis Test were sent.

Bulletin 4 again dealt with reading. It reported to the students the results of the Jones Vocabulary Test which they had been asked to give in Bulletin 1. This bulletin contained a summary of the results, an analysis of these results, suggestions for the use of phonics in the teaching of reading, and basic principles underlying the teaching of phonics. Aside from the work in phonics, reference was made to Bulletin 1 for suggestions as to other reading exercises to be given.

Bulletin 5 is the second bulletin devoted to spelling. While the exercises suggested in Bulletin 2 were based upon the findings of the Morrison McCall Test, the exercises in Bulletin 5 included also the results of the class work carried out in compliance with Bulletin 2. A summary of the returns from all of the classes was given, indicating the number of times each word of the informal test had been misspelled in the first and the second times the test had been taken. Again, as in Bulletin 2, specific spelling difficulties were analyzed and measures to correct such difficulties were recommended.

Bulletin 6. Until the close of the fifth week, each bulletin was devoted to one subject only—arithmetic calculation, spelling, or reading. With the beginning of the sixth week, this policy was changed. Reading and spelling were to be continued, and corrective work in arithmetic problem solution was presented for the first time in Bulletin 6. The work in

arithmetic problem solution was based upon the analysis of the results of The Buckingham Scale for Problem Solution, a summary of which analysis was included in the bulletin.

Bulletins 7 and 8. At this point, the student was left more to his own initiative in organizing his work, reference being made to bulletins sent previously from which he was to draw his material. Bulletins 7 and 8 gave general directions for conducting the work in the four subjects.

Bulletin 9 outlined the testing program planned as a check on the work of the preceding eight weeks.

Bulletins 1 and 5 are offered as samples of the work summarized above.[1]

BULLETIN 1

TEST-DETERMINED INSTRUCTION FOR EXPERIMENTAL PRACTICE CENTERS—GRADES 3, 4, 5

Subject—Reading Number of periods 2
Time—Week of Jan. 5, 1925 Length of periods 30'

Problems for week:

1. To rouse desire for improvement in reading.
2. To make note of individual undesirable reading habits.
3. To give Jones Vocabulary Test to children whose scores in the Stanford Vocabulary Test are below the class average. This test will help determine the specific vocabulary difficulties of these children and, thus, the types of vocabulary exercise to give them.
4. To give one remedial exercise to group in need of work in sentence meaning and one remedial exercise to group in need of work in paragraph meaning.

Organization of the work:

1st period.

Activity 1. Rousing desire for improvement in reading. Present chart to children showing the distribution of their class scores. Let the children find their scores on the chart. Tell of the tests to be given in June. Emphasize the importance of the amount of growth made between the two tests rather than of the scores themselves.

Activity 2. Study of undesirable reading habits. Assign a silent reading lesson to the class as a whole. Observe carefully the children who are below average in Monroe Reading (Comprehension) and in Vocabulary (see accompanying class list). Make note of the following undesirable reading habits:

a. Pointing to words with fingers.
b. Lip movement.
c. Audible repetition of words.
d. Screwing up face, shaking head, tugging at hair.

Make note of the following undesirable study habits:

a. Lack of concentration.
b. Failure to get to work at once.
c. Short duration of close attention.

[1] All other bulletins are on file in the library of The Johns Hopkins University.

The making of these observations will be facilitated if a seating chart of the class is prepared beforehand, the jottings being made immediately upon the chart.

Activity 3. Giving of the Jones Vocabulary Test to children with poor vocabulary scores. While one student teacher is conducting Activity 2, the other two student teachers will give the Jones Vocabulary Test in accordance with directions.

Summary:

1. Presenting chart—Class scores.
2. Children read silently. Student 1 makes observation while Students 2 and 3 give vocabulary test to children individually.

2nd period.

Activity 1. Continuation of Jones Vocabulary Test.

Activity 2. One student will give an exercise in sentence meaning and in paragraph meaning. The group weak in vocabulary and the group not in need of remedial work will work at desks at an assigned reading exercise.

Activity 3. One student will give an exercise in paragraph meaning, the organization of the class being the same as for Activity 2.

Summary:

Directed	Undirected
1. Testing Individuals	
2. Sentence Meaning	Vocabulary Group
3. Paragraph Meaning	Non-Remedial Group

Suggested Exercises for Remedial Work in Reading

The exercises given below are suggested as types from which selection may be made in carrying out the instructions in the bulletin.

A. Exercises in Sentence Meaning

1. *Yes* and *no* questions
   - Lewis and Roland, 3rd Reader, pp. 27, 245
   - Lewis and Roland, 4th Reader, p. 70
   - Lewis and Roland, 5th Reader, p. 125
2. Mixed Sentences
   - Lewis and Roland, 5th Reader, p. 212
3. Completions
   - Maryland School Bulletin, March 1924, vol. II, No. 11, p. 20
   - Elson Book, 3, p. 137
   - Progressive Road to Reading, Book 2, p. 97
4. Choosing the right word
   - Lewis and Roland, Book 3, p. 69
5. Choosing the right answer
   - Lewis and Roland, Book 3, p. 97
6. Following directions
   - Lewis and Roland, Book 3, p. 161
   - Lewis and Roland, Book 4, p. 174
   - Lewis and Roland, Book 5, p. 106
7. True or false statements
   - Lewis and Roland, Book 3, p. 179
   - Lewis and Roland, Book 4, pp. 99, 106

B. Exercises in Paragraph Meaning

1. Acting a paragraph
   Lewis and Roland, Book 3, pp. 60, 204, 130, 199
   Lewis and Roland, Book 4, pp. 148, 25.
2. Three types suggested in Md. School Bulletin, March, 1924, p. 35.

a. "Children read silently one paragraph of the lesson and then glance up. The teacher asks a pivotal question about the paragraph. This is continued throughout the selection." (See also Every Day Classics, 5th, pp. 278-284.)

b. "The teacher explains before the paragraph is read that the purpose of reading is to find the answers quickly. She then asks a question calling for some specific fact. As soon as a pupil can answer the question he raises his hand, or glances up at the teacher. This speeds up the slow reader."

c. "The teacher asks the question before the paragraph is read. Whoever finds the answer may stand and read aloud the one sentence which contains the answer."

3. Let children read a paragraph and find the sentence that the paragraph tells about (Topic Sentence).
4. Teacher writes sentence on board and asks children to guess what paragraph is about. Children then are directed to a certain page and read the paragraph in order to determine how near right their guess was.

C. Exercises for Children Not in Need of Remedial Work

It is important that these children should be given such specific training as will lead them further in their reading appreciation and power. Give them definite assignments and check the results.

1. Assignment in which definite questions on paragraph meaning are asked, when the answer is not found directly in the material but must be inferred.
2. Working up individual reports on supplementary material in history, geography, or nature study and presenting such reports at the class recitations in those subjects.
3. Preparation by the group of a dramatization to present to the class.
4. *The following material is recommended:*

a. Wheeler.
b. Horn and Shields.
c. Elson.

D. Exercises in Vocabulary

1. Forming a definite number of small words from a large word; e. g., Form 12 words from ***Baltimore***.
2. Answering questions by selecting appropriate words.
   Maryland School Bulletin, March, 1924, p. 53.
   Bolenius, Fifth Reader, p. 122.
3. Word pictures.
   Md. School Bulletin, March, 1924, p. 56.
4. Practice in analysis of prefix, suffix, and root.
   Md. School Bulletin, March, 1924, p. 55.

BULLETIN 5

TEST-DETERMINED INSTRUCTION FOR EXPERIMENTAL PRACTICE CENTERS—GRADES 3, 4, 5

Subject—Spelling Time—Week of February 9—City
Week of February 16—County

The experimenter wishes to express her appreciation of the very fine way in which returns from the remedial teaching are being sent in. They are good both in content and in form. The latter is of the utmost importance in this work, for such attention as you are giving to little details facilitates the work of tabulation. Remember, please, to identify clearly every paper that you send in: name of school, grade, date, your name.

Spelling

Time—3 15′ periods

I. Bases of work for week:

1. Returns from informal tests that you have sent in.
2. Needs shown in Morrison McCall Spelling Test.

II. Problems for week:

1. To strive for at least 5% greater accuracy in the words listed than was attained during the previous week.
2. To provide opportunity for children who already know these words to master more difficult ones.

III. Returns from Informal Tests:

The tables below show the number of errors made in the first and second informal tests given in the 6 city and the 6 rural experimental classes—for a total of 225 children:

Grade 3			Grades 4 and 5		
Words	Errors 1st test	Errors 2nd test	Words	Errors 1st test	Errors 2nd test
1. tea	15	0	1. led	97	2
2. red	5	0	2. red	17	1
3. heat	24	3	3. heat	12	2
4. sea	9	3	4. sea	9	2
5. and	12	1	5. and	6	2
6. bead	17	3	6. fled	61	2
7. play	21	1	7. bled	99	4
8. clean	18	4	8. able	35	3
9. above	33	15	9. baby	20	4
10. baby	15	11	10. their	83	7
11. about	29	14	11. hear	33	6
12. flea	66	9	12. know	40	5
13. dear	15	1	13. Ned	69	5
14. cage	24	7	14. shed	49	17
15. damp	39	7	15. wear	68	11
16. each	29	7	16. bed	12	2
17. daily	58	7	17. air	15	3
18. beat	27	5	18. above	29	6
19. child	27	6	19. about	27	5
20. appear	75	10	20. add	6	4
Total	558	114		787	93

What the Results Show:

1. They show a fine gain as the result of three 15′ periods of spelling, a gain of approximately 80% in the 3rd grade and of 88% in the 4th and 5th grades.

2. That there are certain words of peculiar difficulty: Grade 3—above, about, baby, appear; Grade 4—shed, wear, their, hear, above. Can you find out why these words are so difficult?

If you can, include your observations under *Comments* in your diary.

IV. Needs shown in the Morrison McCall Spelling Test.

During your first week of remedial instruction in spelling you saw that the first word in the Morrison McCall Test that presented difficulty in Grades 4 and 5 was the word *led;* in Grade 3—the word *sea.* These are words of the homonym type, and drill lessons were planned to suit the type of difficulty involved. The next most frequently misspelled word in Grade 3 was *mine* and in Grades 4 and 5, *omit.* Lesson types will be given suitable for teaching these words.

V. Lesson Types:

Word: *Mine.*

Difficulty: Confusion with *mind* due to faulty listening and to faulty pronunciation.

Teaching emphasis: Upon listening and pronunciation.

Steps:

1. Meaning. Teacher writes word on board, pronounces it, and calls for sentences.

2. Pronunciation. Teacher pronounces word distinctly and calls attention to the closing sound. Children pronounce individually. Teacher calls for other words that end like *mine.* Teacher writes words on board as children give them. After a number of words have been given, children practice clear pronunciation of them.

3. Children write each word, pronouncing it as they write.

Word: *Omit.*

Difficulty: Confusion of first letter with *a* due to faulty listening and pronunciation; lack of familiarity with meaning of word.

Teaching emphasis: Upon meaning and use; upon listening and pronunciation.

Steps: Same as with word *mine,* emphasis in pronunciation being placed upon the *o.*

VI. Organization of the work: 3-15′ periods.

1st period.

1. Teacher has made a summary of the results of the two informal tests for her class similar to the summary the experimenter has made for all the children. For this purpose your score sheets have been returned to you. Keep these sheets as part of your permanent data. Teacher has written the summary on the board. She calls the children's attention:

a. To the percentage of improvement. (To find this, subtract the number of errors made in the second test from the number of errors made in the first test; divide the remainder by the number of errors made in the first test.) Set the goal of a percentage of improvement 5 points ahead of that made the last time.

b. To the words most frequently misspelled. (These may or may not correspond to the words most frequently misspelled as shown in the summary in this bulletin.) Emphasize the need for mastery of those words.

2. Drill the words most frequently misspelled.

3. Give the informal test for the week. Test 2a in Grade 3 and Test 2b in Grades 4 and 5.

2nd Period:

The teacher has scored the informal tests and made entries on one of the score sheets provided. Group the children according to the results, putting those who made not more than 2 errors in Group 1 and all the rest in Group 2.

	Directed	Undirected
1st 5'.	Groups 1 and 2. Teach the words to all the children which are needed by Group 1.	
Remaining time.	Teach Group 2 as many of the other words misspelled as possible.	Group 1—one of the undirected exercises given in Bulletin 2.

3rd Period:

	Directed	Undirected
1st 10'.	Group 2. Continue teaching words needed.	Group 1. Correct exercises written during last lesson. (A student teacher helps children individually.) Give another undirected exercise, Bulletin 2.

Informal Test—Spelling 2A for Grade 3

1. above	Fayette St. is above Baltimore St.	above
2. mine	The book is mine.	mine
3. appear	The sun will soon appear.	appear
4. line	Draw a straight line.	line
5. about	Let us walk about the garden.	about
6. pine	The pine tree has long needles.	pine
7. baby	The baby is playing.	baby
8. fine	The weather is fine.	fine
9. across	Walk across the room.	across
10. back	Please come back.	back
11. call	Please call me early.	call
12. dance	We are going to a dance.	dance
13. each	Each child has a book.	each
14. fall	Try not to fall.	fall
15. game	We are playing a good game.	game
16. hand	His hand was cold.	hand
17. ice	There is ice on the ground.	ice
18. jump	How high can you jump?	jump
19. keep	Keep your pencils in a box.	keep
20. lake	The lake is deep.	lake

Informal Test—Spelling 2B Grades 4 and 5

1. shed	The wagon is in the shed.	shed
2. omit	Omit the last paragraph.	omit
3. wear	She will wear her new dress.	wear
4. obey	Obey the rules of the game.	obey
5. open	Please open the door.	open

6. their	These are their books.	their
7. hear	I hear the birds singing.	hear
8. above	Fayette St. is above Baltimore St.	above
9. mine	The book is mine.	mine
10. across	Walk across the room.	across
11. back	Please come back.	back
12. call	Please call me early.	call
13. dance	We are going to a dance.	dance
14. each	Each child has a book.	each
15. fall	Try not to fall.	fall
16. game	We are playing a good game.	game
17. hand	His hand was cold.	hand
18. ice	There is ice on the ground.	ice
19. jump	How high can you jump?	jump
20. keep	Keep your pencils in a book.	keep

NOTE.—The above list consists:

1. Of words most frequently misspelled the second time the first informal test was given.
2. Words like *omit* or *mine* frequently misspelled in the Morrison McCall Test.
3. Words selected from the list compiled by Homer J. Smith from the examination of 75,000 running words of the spontaneous compositions of children.

3. EXPERIMENT DURING THE THIRD TERM

While the plan during the third term followed, in general, that of the second term, it was possible to organize the work with greater definiteness than during the second term and to have all subjects taught from the beginning instead of having but one of the four experimental subjects taught during each of the first four weeks as in the second term. The work was planned as follows:

The 51 school days of the term were divided into 3 units of 17 days each. Each of the 3 students assigned to an experimental class carried one of the experimental subjects during a unit of 17 days and shifted to another subject at the end of that unit, arithmetic calculation and problem solution being considered as one subject.

1st unit (17 days)—	Student	A	taught	arithmetic
	"	B	"	reading
	"	C	"	spelling
2nd unit (17 days)—	"	A	"	reading
	"	B	"	spelling
	"	C	"	arithmetic
3rd unit (17 days)—	"	A	"	spelling
	"	B	"	arithmetic
	"	C	"	reading

In all other respects, the experiment was conducted as before. Bulletins were sent out weekly and diary blanks and results were sent to the experimenter. Summaries of the bulletins are offered here.

Bulletin 10 gave specific directions for:

1. Giving the Gray Oral Reading Test
2. Correcting certain type difficulties in spelling
3. Practicing 5, 7, and 9 addends single column addition
4. Use of certain type problems in problem solution

Bulletin 11 reported the results of the tests given at the close of Term 2. In the case of reading and spelling, comparison was made with the results of the fall tests. In addition, detailed score sheets showing the results for each child were sent to each center.

Bulletin 12 summarized the analysis of the March reading tests by classes and individuals, detailed analysis for each class being sent with the bulletin. Special attention was called to the children who had failed to make satisfactory progress in reading, individual diagnoses being sent to the centers. Effort was made to meet the individual needs more effectively by adopting at this point the individual reading system described in Bulletin 12.

In addition to the analysis of the reading situation, work in spelling and arithmetic calculation was directed to be carried on in the usual way, while arithmetic problem work was omitted.

Bulletin 13 called attention to supplementary reading material available for use in the experimental classes. The bulletin also diagnosed the March spelling results in the same way that Bulletin 12 diagnosed the reading results, detailed score sheets being sent to the various centers. Spelling and arithmetic calculation were carried on in the usual way, while problem work was again omitted.

Bulletin 14 gave the details of the plan for reading adopted as a result of the findings of the March tests. Work in spelling, arithmetic calculation, and arithmetic problem solution was also provided.

Bulletin 15 referred to Bulletin 14 for reading aids. It suggested different spelling exercises for the sake of variety

and also gave additional work in arithmetic calculation and arithmetic problem solution.

Bulletin 16 gave directions for a mid-term testing program to be given as a check on the work carried on thus far.

Bulletin 17 carried forward the work in the four subjects in ways outlined previously.

Bulletin 18 reported the results of the April tests in spelling and reading and analyzed the results. In addition, individual score sheets were sent to the centers as they had been sent in March.

Bulletin 19 gave the results of the April tests in arithmetic. No new work was outlined, the students being advised to continue along lines previously indicated.

All bulletins and all forms, record sheets, and score sheets referred to are on file in the library of The Johns Hopkins University. Bulletin 14 is given here in full as a sample of the bulletins used during this term.

BULLETIN 14

Week of April 13, 1925.

TEST-DETERMINED INSTRUCTION FOR EXPERIMENTAL CENTERS

Reading

Foreword to the Instructors:

Our recent testing has shown a fine average gain made by most of our classes. On the other hand, it has shown two conditions which are not gratifying:

1. The percentage of children making no progress is unduly large.
2. The children who are making least progress are those who, in general, made the highest scores in the fall testing.

Our problem has, therefore, become: How can we administer our test-determined instruction so that every child may be insured some growth by reason of that instruction?

If we solve this problem or, even, if we make some advance toward its solution, our next test results should show practically all of the children—good readers and poor readers—pushing ahead of their previous record. The average class gain may be no greater. It may not be as great. That is of little significance. The true criterion of the efficiency of instruction is individual progress proportionate to ability. Can we find a solution?

Let us look at the difficulties:

1. To insure such progress, the instruction must be differentiated.
2. Differentiated instruction involves the division of the class into small groups, or, individual instruction, or, a combination of both.
3. Such instruction is difficult to administer in large classes under ordinary circumstances.

4. The administrative difficulties are increased many times when the work is done by student teachers who—no matter how great their teaching ability is—lack that skill in class management which only experience gives.

In the light of the above difficulties, our major problem includes two specific problems in "ways and means." We must find a way:

1. To reduce the time required for preparation on part of the teacher to a minimum.
2. To work out a scheme of class management that may be reduced to a more easy routine.

The plan submitted in this bulletin is an attempt at a solution of the major problem. At the same time, every possible consideration has been given to the two specific problems. Its success depends, primarily, upon two things:

1. The enthusiasm that the teacher can rouse in her children.
2. The extent to which the teacher can routinize the filing of materials, their distribution and collection, and the keeping of records.

The Plan

The plan provides:

1. Intensive reading of a small unit of material carefully selected with reference to levels of difficulty, such reading to be done under time pressure.
2. Extensive reading of larger units of material carefully selected with reference to levels of difficulty, such reading to be done without time pressure.
3. Drill in mechanics as need for such drill is shown in the two exercises outlined above.

How to carry out the plan:

1. The intensive reading

Materials needed:

Mimeographed exercises; mimeographed question blanks; class record sheet; pupils' record sheets; reading diary blanks.

Plan:

See directions on materials accompanying bulletin.[1]

2. The Extensive Reading

Materials needed:

Books of varying levels of difficulty to suit the needs of the various groups. It is not necessary to have enough of any one kind for a group. Suitability to the needs of the group is the important point.

Class Record. (Same as for intensive reading.) Pupil record for extensive reading. (After Winnetka Plan.)

What to do at first lesson:

a. Explain the plan to the children as follows: That books will be provided for them to be enjoyed as much as possible; that when they have finished reading a book, they are to fill out the book record (show it and explain how to use it): that if the record satisfies the teacher, another book will be given to them.

b. Distribute the books—if you have not already done so—and let the children read silently.

What to do at succeeding lessons:

a. During the first and second periods of each week, the children are to be permitted to go on with their extensive reading after they have corrected their intensive exercises.

---

[1] Appendix K.

b. The third period of each week is to be devoted entirely to this phase of the work.

During the first fifteen minutes, pass among the children giving such help as they seem to need. As in the intensive reading, you will learn gradually how to make this help effective. Let your help always be of such character that it will make the children better able to cope with other difficulties. Rely as much as possible on word analysis in meeting difficulties with words. (See Bulletin 4.)

How much time is to be used:

One full period a week.

Such time as the children have left over from the correction of the exercises during the other two test-determined instruction periods.

Any other free time that the children care to use.

How the work is to be recorded:

a. By the teacher in the column marked for the purpose on the Class Reading Record, such record to be made when a reading unit has been completed.

b. By the pupil when he completes a reading unit.

c. By the teacher on the reading diary blank.

3. Drill

Drill during the test-determined instruction periods is to be confined to such needs as are noted during the other exercises. The experimenter will be able to give definite help in this phase of the work after you send in returns.

Returns to the Experimenter

1. Reading diary blanks.
2. The corrected question blanks.

Additional Notes on Reading

Instructions to Children at Beginning of an Intensive Exercise.

"The exercise which I am about to give you will do much to improve your reading. We will all read for exactly the same length of time. Begin to read when I say 'Read. Stop when I say Stop.' Read as fast as you can, but be sure to get the meaning. The aim is to get all the meaning you can the very first time you read something. After you finish reading, I will give you some questions about what you have read. Let us see how many of the questions you can answer correctly."

Directions for Scoring Intensive Exercises.

The key provided with the exercises gives the answers to the questions expressed in the fewest possible words which may be accepted as correct. The children's answers must include at least all of these words. If additional words are given, consider the answer correct if such words clarify or enrich the meaning. Consider such answers incorrect if the additional words distort the meaning.

Mark each answer 1. The total score is the number of correct answers. Do not mark the incorrect answers. Use a colored pencil in marking.

NOTE.—Carefully file the exercises after the children have used them. What you use for one group one week may be material for other children later. The experimenter will try to gauge from the records and from your reports which exercises you need and will send additional exercises in case of emergency needs. Be as explicit and prompt in your returns as possible, for it is through your returns that the experimenter knows best what to send you.

Spelling

The gratifying results of the second testing in spelling warrant our continuing the same procedure. You are asked to pay particular attention to the following points:

1. Keep the ideal of increase in accuracy before the children constantly. Show them the increase in the average class score in the second test over the first. If a child fails to improve in the second test over the first, try to discover the couse of such failure.

2. Provide worth-while exercise for the group who show, in the first test, that they do not need drill. For suggestions, see Bulletins 2, 5, and 12.

Instead of the diary forms formerly used, please send the experimenter the following data each week:

1. List of the class with the scores made in each of the two tests.
2. The three most frequently misspelled words with the number of errors made in each test.
3. A list of the words that you taught.
4. The type of exercise given the group that did not need the drill.
5. Any comments that you care to make.

Work of the Week:

The work is based upon the word, *friend,* a source of difficulty in the first standard test. The silent *i* is the cause of difficulty.

Word—*friend.*

Difficulty—silent *i.*

Teaching emphasis—visual stimulus and motor response.

Steps:

1. Teacher writes word on board and pronounces it.
2. Children give a few sentences to show that they understand the use of the word.
3. Teacher asks children: "How many sounds are in the word?" (5). How many letters? (6). Which is not sounded?
4. Children write word several times, spelling aloud as they write.

Informal Test

1. friend	She is my friend.	friend
2. girl	She is a pretty girl	girl
3. glass	Please wash the glass.	glass
4. gone	They have gone home.	gone
5. had	She had an interesting book.	had
6. into	He fell into the water.	into
7. thread	The thread is strong.	thread
8. read	I have read the book.	read
9. half	He bought a half pound of candy.	half
10. high	The fence is high.	high
11. just	She has just finished.	just
12. kill	Did he kill the fly?	kill
13. June	School closed in June.	June
14. lake	He rowed on the lake.	lake
15. land	Land was in sight.	land
16. law	He is studying law.	law
17. lot	He bought a house and lot.	lot
18. lay	Lay the book down.	lay
19. led	He led the horse home.	led
20. mine	The book is mine.	mine

Arithmetic Problems

You are asked to organize the work in arithmetic problems similarly to that in spelling. That is:

1. Give 3-15′ periods per week.
2. At the first period, give an informal test to all the children.
3. During the second period, devote the first ten minutes to the children who did not have perfect scores in the test. While you are teaching these children, give more advanced work to the rest. During the last 5 minutes check the work of the latter group. While you are doing this, have the other group correct their papers.
4. During the third period, give another informal test similar to the first one only to those children who failed to have perfect scores in the first. Continue the more advanced undirected work with the other children.

Mark these papers carefully, giving them back from lesson to lesson for correction.

Please remember, that this work emphasizes the reading of the problem and the thinking of the solution. The calculation is purposely kept as simple as possible.

Do not use the old diary form in reporting this work to the experimenter. Instead, send her the following information:

1. A class list showing the scores made by the various children in each test.
2. An example of the type of work that you are giving to the children who do not need this work.

First Informal Test

1. How many are 8 pencils and 4 pencils?
2. If you do 5 examples in arithmetic every day for 5 days, how many have you done in all?
3. If you have 10 pennies and lose 3 of them, how many would you have after you found one of them?
4. How many pies are needed to serve 24 people, if you can cut 6 slices out of a pie?
5. John worked 3 days at $2 a day. Out of his earnings he bought a hat for $2, a book for $1.50, and some candy for $.25. How much has he left?
6. How much change would you get from a dollar bill after buying 1 doz. eggs for $.35, 1 lb. butter for $.50, and 1 loaf of bread for $.10?
7. If the price of lemons is 2 for 5 cts., how many can you buy for $.50?

Second Informal Test

1. How many are 9 cents and 3 cents?
2. If you walk 3 miles a day for 7 days, how many miles have you walked in all?
3. If you have 12 marbles and lose 3 of them and then get 2 others, how many have you?
4. How long will it take you to do 20 problems in arithmetic, if it takes you 5 minutes to do one?
5. John worked 4 days at $1 a day. Out of his earnings he bought a book for $1, a ball for $.75, and some pencils for $.25. How much had he left?

6. How much change will you get from a dollar bill after buying a ball for 50 cents, some marbles for 15 cents, and a ball of cord for 5 cents?

7. If the price of pencils is 2 for 5 cents, how many can you buy for a quarter?

Arithmetic Calculation

Experience has shown that children are inclined to add or to subtract the denominators of common fractions when called on to add or to subtract such fractions. Furthermore, they are apt not to reduce the answers to lowest terms. The drill this week aims to correct these errors and to render automatic the responses to addition and subtraction of simple fractions of the same denominator.

Use any drill device that you wish in doing this work.

If you teach third grade and the children have had no experience with fractions, omit the test-determined calculations entirely this week.

In sending in returns, do not use the old diary form. Instead, send the experimenter the following:

A list of the children's names with scores in the informal test to be given at the third period. Also, add any remarks that you wish.

Informal Test

$\frac{1}{2}$ plus $\frac{1}{2}$	$\frac{1}{3}$ plus $\frac{2}{3}$	$\frac{1}{4}$ plus $\frac{1}{4}$	$\frac{3}{8}$ plus $\frac{1}{8}$	$\frac{2}{5}$ plus $\frac{1}{5}$	
$\frac{3}{4}$ plus $\frac{1}{4}$	$\frac{1}{6}$ plus $\frac{1}{6}$	$\frac{1}{8}$ plus $\frac{5}{8}$	$\frac{5}{8}$ plus $\frac{1}{8}$	$\frac{3}{5}$ plus $\frac{1}{5}$	
$\frac{5}{8} - \frac{1}{8}$	$\frac{3}{8} - \frac{1}{8}$	$\frac{7}{8} - \frac{3}{8}$	$\frac{3}{5} - \frac{1}{5}$	$\frac{4}{5} - \frac{1}{5}$	
$\frac{5}{6} - \frac{1}{6}$	$\frac{3}{4} - \frac{1}{4}$	$\frac{2}{3} - \frac{1}{3}$	$\frac{1}{8} - \frac{1}{8}$	$\frac{1}{4} - \frac{1}{4}$	

CHAPTER VII

The Experiment as it Was Carried Out

To have carried out the work outlined in the bulletins as given in Chapter VI in all the details indicated would have meant the expenditure of time given in Table XII. From

TABLE XII

Expenditure of Time to be Devoted to Experimental Teaching According to Bulletin Directions

Bul.	Reading			Spelling			Arith. Cal.			Arith. Prob.			Total		
	Pd.	L.	T.	Pd.	L.	T.	Pd.	L.	T.	Pd.	L.	T.	Pd.	L.	T.
1	2	30	60										2	30	60
2				3	15	45							3	15	45
3							3	15	45				3	15	45
4	3	30	90										3	30	90
5				3	15	45							3	15	45
6	3	30	90	3	15	45				3	30	90	6	30	180
													3	15	45
7	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
8	3	30	90	3	15	45	3	15	45	3	15	45	9	15	135
9	Testing												3	30	90
Tot. Term 2	14	30	420	15	15	225	9	15	135	3 6	30 15	180	17 30	30 15	510 450*
															960
10	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
11	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
12	3	30	90	3	15	45	3	15	45				3	30	90
													6	15	90
13	3	30	90	3	15	45	3	15	45				3	30	90
													6	15	90
14	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
15	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
16	Testing														
17	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
18	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
19	3	30	90	3	15	45	3	15	45	3	15	45	3	30	90
													9	15	135
Tot. Term 3	27	30	810	27	15	405	27	15	405	21	15	315	27 75	30 15	810 1125
															1935
Tot. Both Terms	41	30	1230	41	15	630	36	15	540	30	15	495	44 105	30 15	1320 1575
															2895 or 48 hr. 15′

* Length of periods expressed in minutes.
Pd.—Number of periods.
L.—Length.
T.—Total.

the viewpoint of the amount of time planned for the experiment, the above table indicates that: 1. More time was given in the third term than in the second; 2. The work was more regularly arranged in the third term than in the second; 3. In both terms more time was given to reading than to any one of the other three subjects.

The amount of time actually devoted to the work in each of the experimental centers was calculated. The source of the information is the diary blanks on which the students were asked to indicate the amount of time actually used. While this information is subject to the usual error found wherever such method is used, all precautions were taken to procure accurate records in this respect.[1]

Tables XIII, XIV, and XV summarize the expenditure

TABLE XIII

TOTAL EXPENDITURE OF TIME REPORTED OF EXPERIMENTAL TEACHING ON EACH SUBJECT IN EACH EXPERIMENTAL CENTER

Reading Allotted time 1230′		Spelling Allotted time 630′		Arith. Cal. Allotted time 540′		Arith. Prob. Allotted time 495′	
H	1290	A	640	H	585	L	460
A	1210	B	545	I	465	F	420
M	1190	F	497	D	465	I	405
G	1115	H	495	L	435	A	405
L	1058	D	480	A	430	M	395
E	1050	JK	450	E	400	D	365
F	1042	C	435	F	394	G	345
I	975	E	430	G	330	H	345
JK	945	G	405	M	265	C	310
C	916	I	385	C	250	B	270
D	870	M	360	B	230	E	255
B	660	L	265	JK	230	JK	180

TABLE XIV

AVERAGE EXPENDITURE OF TIME REPORTED ON EXPERIMENTAL TEACHING OF EACH SUBJECT IN EACH TYPE SCHOOL SITUATION

	City	Campus	Rural
Reading	958.00′	1126.66′	1064.30′
Spelling	504.50′	428.30′	358.20′
Arithmetic Calc.	361.50′	460.00′	309.30′
Arithmetic Prob.	337.50′	365.00′	345.00′

[1] Tables indicating the amount of time reported actually given each week in each school to each subject are filed in the library of The Johns Hopkins University.

TABLE XV

AVERAGE EXPENDITURE OF TIME REPORTED ON EXPERIMENTAL TEACHING OF EACH SUBJECT IN EACH GRADE AND IN ENTIRE EXPERIMENTAL GROUP

	Grade 3	Grade 4	Grade 5	Total
Reading	982.50′	1047.25′	1043.33′	1026.75′
Spelling	517.50′	414.16′	397.83′	448.91′
Arithmetic Calc.	305.00′	371.33′	364.50′	373.08′
Arithmetic Prob.	300.00′	342.50′	352.60′	346.25′

of time used in test-determined instruction in the four subjects by individual classes, type school situation, and by grade respectively.

Table XIII shows no tendency toward constancy of rank in time expenditures by any particular class or classes. Thus, Class H which ranks first in time allotted to reading, ranks fourth in spelling, first in arithmetic calculation, and eighth in arithmetic problem solution. Class B which ranks twelfth in time allotted to reading ranks second in spelling, eleventh in arithmetic calculation, and tenth in arithmetic problem solution.

Table XIV shows that the campus elementary school in general gave more time to the experimental work than the city and rural schools gave. Averaging the ranks of each, the relative positions in time expenditure of the three type school situations are: 1. Campus, 1.25; 2. City, 2.25; 3. Rural, 2.50.

Table XV shows a tendency in Grade 4 to outrank the other two grades in amount of time expended. Averaging the ranks for each grade, the relative positions in time expenditure are: 1. Grade 4, 1.50; 2. Grade 5, 2.00; 3. Grade 3, 2.50.

No doubt the character of the work done by the students varied to as great, if not a greater, extent than the time used. Of this, there is no measure. The character of the diary reports varied from those which gave the experimenter but little information to those done with the utmost care. Three samples of such diary reports are offered as types of poor, average and superior reports.

Sample of a Poor Diary

DIARY: TEST-DETERMINED INSTRUCTIONS

1. School ——— 2. Grade 4 & 5 3. Date Jan. 27 4. Time 10.45-11.00
5. Num. of Minutes 7½ Activity 1: From 10.45 to 10.52
" 2: From to
6. Subject Spelling " 3: From to
8. Types of Instruction 7. Groups or class
Activity 1 Directed
" 2 Undirected
" 3
9. Materials Used:
Bulletin 2
10. Comments: .
Work went very well
11. Disciplinary Cases:
12. Absent:
(Name)
(Name)

Approved: Signed ———
Student Teacher

Sample of a Fair Diary

DIARY: TEST-DETERMINED INSTRUCTIONS

1. School ——— 2. Grade Fifth 3. Date Jan. 19
4. Time Activity 1: From 10.45 to 10.50
" 2: From 10.50 to 11.00
" 3: From to
5. Num. Minutes Fifteen 7. Groups or class:
6. Subject Spelling Class as a whole
8. Types of Instruction:
Activity 1: Presenting chart
" 2: Giving informal test
" 3:
9. Materials Used: Used words given in Bulletin 2. Marked, informal test Grades 4 and 5
10. Comments: There were a few youngsters who remarked, while I was dictating the words, that these words were baby words. The pupils who said this were 3, 6, 8
11. Disciplinary Cases: There was nothing unusual in this lesson regarding discipline. The test was given without any interference from the pupils
12. Absent:
None were absent

Approved: Signed ———
Student Teacher

SAMPLE OF A GOOD DIARY

DIARY: TEST-DETERMINED INSTRUCTION

School	Grade	Date
M. S. E.	4	April 1, 1925

Number of minutes—30
Subject—Reading
Group—Experimental
Types of instruction
Activity 1 Time 1.45-2.15
" 2
" 3
Materials used:
Series of exercises sent out by the experimenter
Comments

The exercises are given out according to the instruction given in Bulletin 12. We have no 4th group to work with. At first the material seemed difficult for all the groups but after they had mastered the first and second problems all seemed very much interested in the work.

Results

Group III handed in very good papers. . . . . had a perfect paper. The results from the others in this group showed that they were capable of doing more difficult work if they would just try to do so.

Group II handed in very good papers. Very few errors were made and these were due to carelessness in reading the paragraph rather than to lack of comprehension of what was read.

Group I seemed to have more difficulty with the exercise given them than the other groups did. Question 1 seemed quite misleading to them. Other errors were due to not thinking through or puzzling out the answer to the question asked.

Disciplinary Cases:
Absent: Signed _____

There were many evidences in the diary reports of careful observation of the children and of study of their problems. A few random samples are quoted:

"Most of the children did not know the meaning of the word fled." "Three of the children said that the words in the test were baby words." "I dictated the words a little too fast in the second test which, I think, accounts for some of the low scores." "Some of the children counted by making dots." "The children wanted to multiply in every problem at first." (The problems included miscellaneous processes.) "In writing the word baby many of the b's were made like l's; in now, the w was made like u." "The children count on their fingers; more drill in tables is needed." "The children do not know how to use 0 in subtraction." "I made a mistake in teaching

the word speech. I spelled it speach." "Children realized their difficulty in 'taking away' in subtraction and asked for help." "The first information tests in spelling were accidentally destroyed and the work in spelling was discontinued for the week." "The time allowed for the reading exercises seems too short to make them of value."

In addition to such comments, many of the diaries included detailed analyses of the work of individual children. On the whole, there was a conscientious effort to carry out the work in accordance with the directions in the bulletin, though the work was subject to many interruptions.

CHAPTER VIII

The Results

1. THE MAIN CRITERION: THE PROGRESS OF THE PUPILS

The results were computed on the basis of the final equating as given in Chapter IV:

1. By the scores on the separate tests.
2. By the composite scores of the tests in reading, in spelling, in arithmetic calculation, and in arithmetic problem solution.

The method of computation and the results of each of these two phases will be treated separately.

The Results as Measured by Difference Between Initial and Final Status in Each Test

Each pupil's score in each initial test was subtracted from his score in the corresponding final test. These differences were averaged for control and experimental groups:

1. By individual classes.
2. By the combination of classes into three experimental and three control groups according to grade—Grades 3, 4, and 5.
3. By the combination of classes into three experimental and three control groups according to type school—city, campus, and rural.
4. By the combination of all classes and schools into one experimental and one control group.

Tabulation of the detailed results is filed in the library of The Johns Hopkins University. A summary of results according to individual classes is shown in Tables XVIa and XVIb and a summary according to grade, type of school, and total experimental and control groups is given in Table XVII.

Comparative gains of the experimental and control groups are given in terms of mean only. (Variabilities and reliabilities are computed for the composites.)

Attempt is made in Figures 11, 12, 13, 14, 15, 16, 17, and 18 to show the relative average gains of the experimental and control groups by comparison with a normal year's growth in the various tests. The normal growth between the beginning and the end of the 3rd grade according to the Stanford

TABLE XVIA

Summary of Average Gains by Classes—Experimental

	A	B	C	D	E	F	G	H	I	JK	L	M	Exper. Gr.*
Reading													
T. McC.	13.57	13.04	13.95	9.61	13.73	5.06	7.16	12.50	8.91	9.82	6.75	11.00	9 classes
S.	29.90	30.56	25.85	21.14	26.86	10.77	20.71	32.75	17.36	11.41	13.25	15.80	9
M. C.	1.16	1.25	2.36	5.17	.82	2.55	.50	.20	.25	1.06	3.16	−1.22	9
												Total	27
Spelling													
M. McC.	6.33	7.22	3.55	5.11	2.31	3.22	5.80	1.88	3.58	4.33	2.00	1.88	7
S.	19.03	14.11	10.48	11.85	10.12	7.00	30.75	22.22	15.27	9.66	4.66	16.33	7
												Total	14
Arithmetic Calculation													
M.	12.39	16.42	9.01	24.24	11.73	8.60	6.51	6.67	9.64	12.38	11.16	17.98	9
S.	26.75	31.65	8.00	26.51	13.52	13.26	27.33	21.60	13.38	19.69	10.28	18.44	6
C.	24.53	50.10	14.21	41.82	−2.82	6.72	23.66	21.33	12.50	18.00	19.00	18.16	9
												Total	24
Arithmetic Problem Solution													
B.	5.13	8.90	4.52	3.35	5.81	4.76	7.50	3.00	9.00	5.42	1.75	3.57	6
S.	1.08	9.00	7.86	2.36	8.40	5.24	5.33	14.66	12.00	8.80	9.33	9.100	5
												Total	11
												Grand Total	76

T. McC. Thorndike McCall Reading
S. Stanford Achievement
M. C. Monroe Reading Comprehension
C. Courtis

M. McC. Morrison McCall Spelling
M. Monroe
B. Buckingham

* Experimental Greater

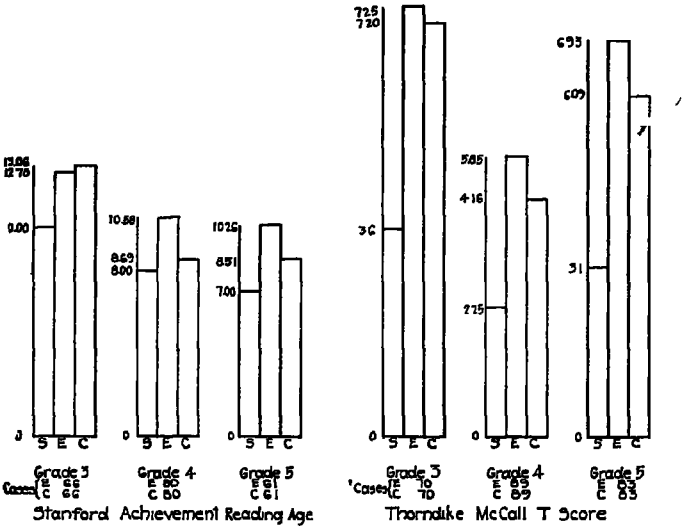

FIGURE 11.—Comparison of gains made by experimental and control classes ing. S: A normal year's growth; E: Experimental; C: Control. The graphs arranged in the same order.

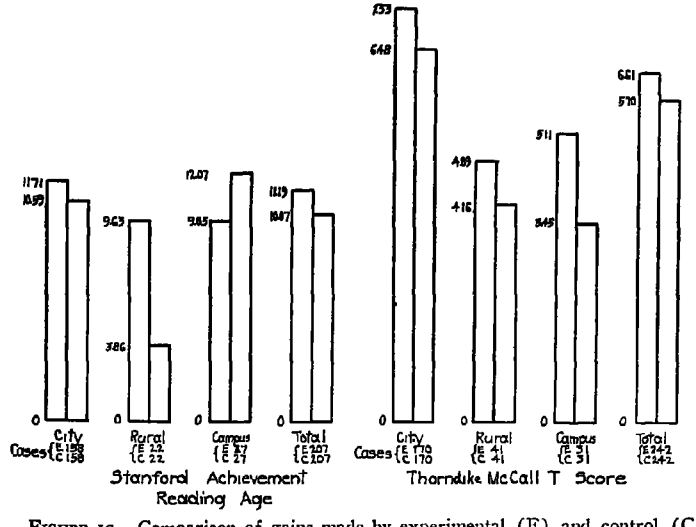

FIGURE 12.—Comparison of gains made by experimental (E) and control (C) according to type school situations: City, rural, ca

TABLE XVIb

SUMMARY OF AVERAGE GAINS BY CLASSES—CONTROL

	A	B	C	D	E	F	G	H	I	JK	L	M	Control Greater
Reading													
T. McC.	11.80	16.82	9.76	8.96	10.66	4.11	7.50	11.70	14.91	1.27	2.50	8.71	3
S.	27.39	33.60	21.61	24.00	21.20	8.61	16.57	27.00	28.36	10.58	5.25	15.50	3
M. C.	.70	1.00	−2.81	.82	.78	.05	−.16	1.80	1.25	.62	.83	−.55	3
												Total	9
Spelling													
M. McC.	4.40	6.04	4.44	5.26	5.38	3.12	5.40	2.66	2.66	2.91	3.75	1.82	5
S.	16.41	11.52	3.04	14.50	17.88	9.20	24.25	14.22	14.36	11.00	−3.33	18.33	5
												Total	10
Arithmetic Calculation													
M.	11.76	13.88	8.92	19.33	11.88	6.86	8.55	6.23	9.39	9.30	16.55	12.51	3
S.	27.31	28.17	4.90	19.54	10.00	19.58	29.33	8.40	14.46	23.23	18.85	12.88	6
C.	18.26	36.80	5.57	37.73	−14.89	5.16	25.88	20.75	12.00	29.12	32.80	16.67	3
												Total	12
Arithmetic Problem Solution													
B.	6.44	6.66	4.12	5.87	7.18	3.96	6.37	−.88	5.07	8.42	8.50	5.80	6
S.	1.04	12.20	8.13	5.18	3.30	6.30	3.33	6.00	15.27	8.53	9.33	10.00	6
												Total	12
												Grand Total	43

TABLE XVII

SUMMARY OF AVERAGE GAINS BY GRADES, TYPE SCHOOLS AND TOTALS[1]

	Total City	Total Campus	Total Rural	Total Gr. 3	Total Gr. 4	Total Gr. 5	Total All
EXPERIMENTAL							
Reading							
T. McC.	7.33	5.11	4.89	7.25	5.81	6.93	6.61
S.	11.71	9.85	9.63	12.78	10.58	10.26	11.19
M. C.	1.17	.31	.80	1.10	.71	1.22	1.03
Spelling							
M. McC.	4.31	3.80	2.78	6.02	3.51	2.85	4.03
S.	12.17	21.17	12.16	18.84	12.83	10.23	13.43
Arithmetic Calculation							
M.	14.50	7.96	14.49	12.72	15.75	11.28	13.25
S.	20.29	19.26	17.36	28.90	14.10	18.73	19.67
C.	23.02	18.76	18.21	31.53	29.05	4.96	21.42
Arithmetic Problems							
B.	5.27	6.80	4.16	6.33	4.14	5.73	5.31
S.	7.30	11.72	9.02	9.20	7.81	7.80	8.18
CONTROL							
Reading							
T. McC.	6.48	3.45	4.16	7.20	4.16	6.09	5.70
S.	10.59	12.07	3.86	13.06	8.69	8.51	10.07
M. C.	.15	.94	.03	.75	−.41	.43	.21
Spelling							
M. McC.	4.87	3.54	2.45	5.13	4.48	3.62	4.37
S.	12.68	16.55	12.27	16.07	9.86	14.94	13.19
Arithmetic Calculation							
M.	12.22	8.56	11.63	11.76	13.17	10.02	11.63
S.	19.09	15.20	18.33	28.51	13.87	15.37	18.33
C.	14.56	18.97	23.84	28.00	23.24	3.27	17.55
Arithmetic Problems							
B.	5.71	3.30	7.31	6.73	4.56	6.08	5.66
S.	7.66	8.96	9.46	10.03	7.21	7.65	8.12

TABLE XVII.—*Continued*

	Type School		Grades		Total	
	Greater E	Greater C	Greater E	Greater C	Greater E	Greater C
			Reading			
T. McC.	3	0	3	0	1	
S.	2	1	2	1	1	
M. C.	2	1	3	0	1	
Total	7	2	8	1	3	0
			Spelling			
M. McC.	2	1	1	2		1
S.	1	2	2	1	1	
Total	3	3	3	3	1	1
			Arithmetic Calculation			
M.	2	1	3	0	1	
S.	2	1	3	0	1	
C.	1	2	3	0	1	
Total	5	4	9	0	3	0
			Arithmetic Problems			
B.	1	2	0	3		1
S.	1	2	1	2	1	
Total	2	4	1	5	1	1
	17	13	21	9	8	2

[1] The initials appearing at the left of the table designate the various tests used, as follows:

T. McC.	Thorndike McCall Reading
S.	Stanford Achievement—Reading—Spelling—Arithmetic
M. C.	Monroe Reading Comprehension
M. McC.	Morrison McCall Spelling
M.	Monroe
C.	Courtis
B.	Buckingham

Reading Age norms was taken as the base (Figure 11, 1st column). The growth made by the experimental and control classes was measured proportionately. Similarly, Grades 4 and 5 were blocked off. Growth as measured by the other tests was blocked proportionately also to the Stanford norm. Thus, in Thorndike McCall Reading, a growth in terms of T Score of 3.6 points is normal for the time between the beginning and the end of the 3rd grade. Hence, a column of

the same height as the column for normal third grade growth in the Stanford Test was used to indicate the 3.6 points of growth according to Thorndike McCall, while the actual growth made by the experimental and control classes was measured proportionately.

The tables and figures indicate:

1. That when the results are studied by individual classes (Table XVIa, b), more experimental than control classes make greater gain in reading, arithmetic calculation, and spelling, while in arithmetic problem solution the balance is slightly in favor of the control group.

27 experimental and 9 control classes make greater average gain in reading.

24 experimental and 12 control classes make greater average gain in arithmetic calculation.

14 experimental and 10 control classes make greater average gain in spelling.

11 experimental and 12 control classes make greater average gain in arithmetic problem solution, while in one case the average gain of the experimental and the control groups is the same.

2. That when the classes are combined by grades (Table XVII), in every test in each grade in arithmetic calculation and in every case but one in reading, the experimental groups make the greater gain; while in spelling the classes making the greater gain are evenly divided between the experimental and control groups, and in arithmetic problem solution the balance is in favor of the control group.

8 experimental and 1 control exceed—reading.
9 experimental and 0 control exceed—arithmetic calc.
3 experimental and 3 control exceed—spelling.
1 experimental and 5 control exceed—arithmetic prob.

3. That when the classes are combined by types of schools in reading and arithmetic calculation (Table XVII), a greater number of experimental classes make the greater gain, while in spelling the numbers are the same, and in problem solution the balance is in favor of the control groups.

4. That the three type school situations differ in respect to relative gains of experimental and control groups in the various subjects as follows:

City:
3 experimental and 0 control exceed—reading.
0 experimental and 2 control exceed—spelling.
3 experimental and 0 control exceed—arithmetic calc.
0 experimental and 2 control exceed—arithmetic prob.

Campus:
1 experimental and 2 control exceed—reading.
2 experimental and 0 control exceed—spelling.
1 experimental and 2 control exceed—arithmetic calc.
2 experimental and 0 control exceed—arithmetic prob.

Rural:
3 experimental and 0 control exceed—reading.
1 experimental and 1 control exceed—spelling.
1 experimental and 2 control exceed—arithmetic calc.
0 experimental and 2 control exceed—arithmetic prob.

5. That among the city and campus classes, a greater number of excess gains is made by the experimental than by the control, while in the rural classes the number of excess gains of experimental and control classes is the same.

City: In 6 tests the average is greater for experimental and in 4 for control.

Campus: In 6 tests the average is greater for experimental and in 4 for control.

Rural: In 5 tests, the average is greater for experimental and in 5 for control.

6. That when all classes are combined regardless of grade or type of school into two big groups—experimental and control (Table XVII)—in each of the three tests in reading and in each of the three tests in arithmetic calculation, the experimental group exceeds the control, while in each of the two tests in problem solution and in spelling, the experimental group exceeds in one and the control in the other.

Reading:	Experimental exceeds in 3 and control in 0 tests.
Ar. Calc.:	Experimental exceeds in 3 and control in 0 tests.
Spelling:	Experimental exceeds in 1 and control in 1 tests.
Ar. Prob.:	Experimental exceeds in 1 and control in 1 tests.

7. That the amount of difference between the gains made by the experimental and control groups is appreciable and varies with the grade and type school (Figures 11, 12, 13, 14, 15, 16, 17, and 18).

The Results as Measured by Difference Between Initial and Final Status in Composite Scores

Composite scores of the results of the final tests were derived in the same way as were the composite scores in the initial tests (Chapter IV). In each case, the difference between the initial and the final composite scores was computed and the scores averaged by grade, type of school, and total group. The results were computed for the composites derived from both the weighted and the unweighted scores as expressed in mean, standard deviation, and reliability.

The reliability of the gains; i. e., the reliability as influenced by the number of cases, was computed according to the formula for Standard Error (S. E.):

$$\text{S. E.} = \frac{\text{S. D. dis.}}{\sqrt{\text{N}}}$$

The results in terms of mean, S. D., and S. E. are given in Tables XVIII, XIX, XX, XXI, XXII, XXIII, XXIV, and XXV.

TABLE XVIII

COMPOSITE READING GAINS BY GRADES

Grade	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
			UNWEIGHTED					
3	29	49.24	15.40	9.15	29	55.06	18.10	3.36
4	39	40.66	15.40	2.46	39	27.74	20.70	3.30
5	33	23.54	21.50	3.74	33	28.36	20.40	3.55
Total	101	37.53	20.30	2.02	101	35.79	22.90	2.27
			WEIGHTED					
3	29	108.00	39.40	7.32	29	114.89	42.40	7.88
4	39	92.46	54.20	8.68	39	68.48	55.60	8.91
5	33	58.09	47.60	8.29	33	63.82	46.80	8.15
Total	101	85.69	55.40	5.51	101	80.28	54.80	5.45

TABLE XIX

COMPOSITE SPELLING GAINS BY GRADES

Grade	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
			UNWEIGHTED					
3	35	37.63	22.30	3.77	35	37.23	19.70	3.33
4	51	22.06	19.30	2.70	51	22.41	16.60	2.32
5	43	18.42	17.90	2.73	43	25.63	19.30	2.94
Total	129	25.04	21.60	1.84	129	27.50	19.40	1.70
			WEIGHTED					
3	33	55.65	33.70	5.70	35	52.94	27.10	4.58
4	51	29.59	28.90	4.04	51	31.72	25.60	3.58
5	43	29.44	32.80	4.04	43	38.42	28.90	4.41
Total	129	36.61	31.50	2.79	129	39.69	28.50	2.50

TABLE XX

COMPOSITE ARITHMETIC CALCULATION GAINS BY GRADES

Grade	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
			UNWEIGHTED					
3	27	96.29	30.90	5.95	27	81.26	32.90	6.34
4	40	43.50	33.00	5.22	40	36.67	43.00	6.96
5	37	54.92	25.00	1.82	37	46.32	27.90	4.58
Total	104	61.27	36.70	3.60	104	51.69	36.70	3.60
			WEIGHTED					
3	27	125.21	38.90	7.49	27	113.66	43.40	4.18
4	40	56.72	42.20	6.67	40	52.25	23.00	3.63
5	37	73.97	41.90	6.89	37	59.05	37.40	6.15
Total	104	80.64	50.00	4.97	104	70.61	49.80	4.95

TABLE XXI

COMPOSITE ARITHMETIC PROBLEM GAINS BY GRADES

Grade	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
			UNWEIGHTED					
3	33	15.78	16.60	2.87	33	19.51	11.00	1.90
4	57	11.49	20.50	2.71	57	7.58	14.90	1.97
5	55	5.50	16.40	2.21	55	4.50	17.50	2.36
Total	145	10.71	18.70	1.55	145	9.80	22.20	1.89
			WEIGHTED					
3	33	23.79	24.20	4.19	33	28.82	21.80	3.77
4	57	18.22	27.00	3.58	57	14.65	26.90	3.56
5	55	16.32	27.60	3.72	55	11.43	29.20	3.94
Total	145	15.74	27.00	2.24	145	13.87	27.70	2.30

TABLE XXII

COMPOSITE READING GAINS BY TYPE SCHOOLS

Type	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
			UNWEIGHTED					
City	85	38.98	21.20	2.29	81	37.24	22.30	2.48
Campus	10	28.00	26.40	8.35	12	33.58	22.40	6.47
Rural	6	32.83	9.50	3.89	8	24.37	23.70	7.15
			WEIGHTED					
City	85	87.96	56.95	6.17	81	83.39	51.00	5.66
Campus	10	77.30	52.62	16.64	12	84.75	34.33	9.92
Rural	6	67.50	33.89	13.89	8	42.12	65.61	23.23

TABLE XXIII

COMPOSITE SPELLING GAINS BY TYPE SCHOOLS

Type	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
			UNWEIGHTED					
City	98	24.15	19.40	1.95	93	27.50	17.90	1.80
Campus	19	36.47	20.50	4.71	19	35.88	23.90	5.49
Rural	12	17.58	27.80	8.03	17	15.72	18.50	5.34
			WEIGHTED					
City	98	33.88	28.25	2.85	93	40.60	26.85	2.71
Campus	19	55.10	31.30	7.19	19	49.63	27.75	6.38
Rural	12	29.58	43.30	1.25	17	23.76	28.18	9.75

TABLE XXIV

COMPOSITE ARITHMETIC CALCULATION GAINS BY TYPE SCHOOL

Type	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
				UNWEIGHTED				
City	76	65.43	43.20	4.96	76	51.05	38.80	4.45
Campus	19	50.42	25.10	5.77	16	42.00	25.90	6.47
Rural	9	49.00	32.40	10.80	12	69.50	38.20	1.10
				WEIGHTED				
City	76	85.43	52.95	6.07	76	73.59	50.55	5.80
Campus	19	67.15	38.70	8.89	16	48.62	38.70	9.67
Rural	9	68.66	34.60	11.53	12	81.08	47.30	1.36

TABLE XXV

COMPOSITE ARITHMETIC PROBLEM GAINS BY TYPE SCHOOL

Type	Experimental Cases	Mean	S. D.	S. E.	Control Cases	Mean	S. D.	S. E.
				UNWEIGHTED				
City	107	9.22	16.60	1.60	109	10.00	21.12	2.02
Campus	22	15.36	18.00	3.83	18	13.80	19.00	4.48
Rural	16	11.94	30.00	4.48	18	8.20	31.00	7.31
				WEIGHTED				
City	107	16.37	29.34	2.83	109	14.69	26.85	2.57
Campus	22	26.13	23.70	5.05	18	20.89	33.75	7.95
Rural	16	21.18	19.70	4.64	18	24.28	23.20	5.47

The reliability of the difference of the gains, given in Tables XXVI and XXVII, was calculated according to the formula:

$$\text{S. E. diff.} = \sqrt{\text{S. E. av. } 1^2 + \text{S. E. av. } 2^2}$$

TABLE XXVI

COMPOSITE GAINS BY GRADES

Gr.	Unweighted: Cases	*Excess Ex.	Excess Cont.	†S. E.	U. of M.‡	Weighted: Excess Ex.	Excess Cont.	S. E.	Reliability
				READING					
3	29		5.82	9.74	71		6.89	10.75	74
4	39	12.92		4.11	100	23.98		12.43	84
5	33		4.82	5.15	83		5.73	11.63	68
Total	101	1.74		3.03	72	5.41		7.75	69
				ARITHMETIC PROBLEMS					
3	33		3.73	3.44	85		5.09	5.63	82
4	57	3.91		3.35	86	3.55		5.04	76
5	55	1.00		3.23	62	3.89		5.42	76
Total	145	.91		2.44	65	1.87		3.21	72
				ARITHMETIC CALCULATION					
3	27	15.03		8.69	96	11.55		8.52	90
4	40	6.83		8.70	78	4.45		7.59	72
5	37	8.60		4.50	97	14.92		9.23	93
Total	104	9.58		5.09	96	10.03		7.01	85
				SPELLING					
3	35	.60		5.03	55	2.71		7.31	65
4	51		.35	3.56	54		2.13	5.39	65
5	43		7.21	4.01	96		8.98	5.97	92
Total	129		2.46	2.50	83		3.08	3.74	74

* Excess Ex.—Number of points, mean gain of experimental group exceeded that of control.
Excess Cont.—Number of points, mean gain of control group exceeded that of experimental
† S. E.—Standard error of the difference.
‡ U. of M.—Unreliability of the Measure—chances in 100 that the true difference or excess gain is greater than 0.

TABLE XXVII

Composite Gains by Type Schools

Type	Cases	Unweighted *Excess Ex.	Unweighted Excess Cont.	Unweighted †S. E.	Unweighted U. of M.‡	Weighted Excess Ex.	Weighted Excess Cont.	Weighted S. E.	Weighted Reliability
				READING					
City	85-E 81-C	1.24		3.37	64	4.57		10.56	66
Campus	10-E 12-C		5.58	10.56	70		7.45	19.33	65
Rural	6-E 8-C	8.46		8.14	84	25.38		20.80	88
				ARITHMETIC PROBLEMS					
City	107-E 109-C		.78	2.58	62	1.68		3.82	66
Campus	22-E 18-C	1.56		5.89	60	5.24		9.41	71
Rural	16-E 18-C	3.74		8.58	66		3.10	7.18	66
				ARITHMETIC CALCULATION					
City	76-E 76-C	14.38		6.66	98	11.84		8.37	92
Campus	19-E 16-C	8.42		8.66	84	18.53		13.13	92
Rural	9-E 12-C		20.50	10.85	97		12.42	10.59	86
				SPELLING					
City	98-E 93-C		3.35	2.65	88		6.72	3.93	96
Campus	19-E 19-C	.59		7.23	53	5.47		9.59	72
Rural	12-E 17-C	1.86		9.64	58	5.82		9.83	59

* Excess Ex.—Number of points, mean gain of experimental group exceeded that of control.
Excess Cont.—Number of points, mean gain of control group exceeded that of experimental
† S. E.—Standard error of the difference.
‡ U. of M.—Unreliability of the Measure—chances in 100 that the true difference or excess gain is greater than 0.

Comparative mean gains of experimental and control groups are shown by grades in Figures 19, 20, 21, and 22 and by type school in Figures 23, 24, 25, and 26.

The tables and figures show:

1. A general agreement between results in the unweighted and the weighted scores, there being but two exceptions to this, in the city and rural results in arithmetic problems. (Tables XVIII-XXV; Figures 19-26.)

2. That when the pupils are grouped according to grades (Tables XVIII, XIX, XX, and XXI; Figures 19, 20, 21, and 22), the balance is in favor of the experimental grades in arithmetic calculation and in arithmetic problems but in favor of the control grades in reading and spelling, while the experimental grades exceed the control in the greater number of total instances.

Arith. Calc.: Experimental exceeds in 6; control, 0.
Arith. Prob.: Experimental exceeds in 4; control, 2.
Reading: Experimental exceeds in 2; control, 4.
Spelling: Experimental exceeds in 2; control, 4.

3. That when the pupils are grouped according to type school in each subject (Tables XXII, XXIII, XXIV, and XXV; Figures 23, 24, 25, and 26), the number of instances in which the experimental groups exceed the control is greater than the number of instances in which the control groups exceed the experimental.

Arith. Calc.: Experimental exceeds in 4; control, 2.
Arith. Prob.: Experimental exceeds in 4; control, 2.
Reading: Experimental exceeds in 4; control, 2.
Spelling: Experimental exceeds in 4; control, 2.

4. That the experimental group of Grade 3 shows poorest results in arithmetic problem solution and reading; Grade 4 in spelling; Grade 5 in spelling and reading.

5. That the experimental group shows the best results in Grade 4, the number of instances according to grade in which the experimental group exceeds the control being as follows:

Grade 4: 7.
Grade 3: 5.
Grade 5: 4.

6. That the experimental group of the city schools shows poorest results in spelling, the campus school in reading, and the rural schools in arithmetic calculation. The number of instances in which the experimental group exceeds the control in each of the three groups is as follows:

City: 5, experimental; 3, control.
Campus: 6, experimental; 2, control.
Rural: 5, experimental; 3, control.

7. That when the pupils are grouped, regardless of grade and type of school, into one experimental and one control group, the balance is in favor of the experimental group in reading, arithmetic calculation, and arithmetic problem solution, and in favor of the control group in spelling.

8. That in every case the chances are more than even that the results in so far as the number of cases is concerned are reliable, while by far the greatest reliability is found in the results in arithmetic calculation. (Tables XXVI and XXVII.)

Percentage of Pupils in Experimental and Control Groups Exceeding Average Gain of Combined Experimental and Control Groups

The average of the weighted gains of the combined experimental and control groups of each grade in each subject was computed. The number of pupils in the experimental and in the control groups who exceeded the average of the combined groups was noted by grade and subject and the percentage that this number was of the total was computed. The results are shown in Tables XXVIIIa and XXVIIIb.

TABLE XXVIIIa

Pupils of Experimental Group Above and Below Average Gains of Combined Experimental and Control Groups

Gr.	Reading		Arith. Calc.		Arith. Prob.		Spelling		Total all Subjects		% all Subjects	
	*A	†B	A	B	A	B	A	B	A	B	A	B
3	8	21	23	4	12	21	3	32	46	78	37.1	62.9
4	25	14	21	19	15	42	29	22	90	97	48.1	51.9
5	13	20	22	15	13	42	23	20	71	97	42.2	57.8
Total	46	55	66	48	40	105	55	74	207	272	43.2	56.8
%	45.5		57.9		27.5		42.5					

TABLE XXVIIIb

Pupils of Control Group Above and Below Average Gain of Combined Experimental and Control Groups

Gr.	Reading		Arith. Calc.		Arith. Prob.		Spelling		Total all Subjects		% all Subjects	
	A	B	A	B	A	B	A	B	A	B	A	B
3	13	16	22	5	13	20	6	29	54	70	43.5	56.5
4	21	18	20.	20	14	43	30	21	85	102	45.4	54.6
5	13	20	16	21	11	44	22	21	62	106	36.9	63.1
Total	47	54	58	46	38	107	58	71	201	278	41.9	58.1
%	46.5		50.9		26.2		44.9					

* A—Above average of combined experimental and control groups.
† B—Below average of combined experimental and control groups.

The results indicate that in seven instances the experimental pupils exceed the average gain of the combined experimental and control groups of any one grade; in four instances the pupils of the control group exceed; and in one instance the numbers of pupils in experimental and control groups who exceed are equal.

When all grades are combined by subject, the number of experimental pupils exceeding the average gain of combined experimental and control groups is greater in arithmetic cal-

culation and arithmetic problem solution, while the number of control pupils exceeding the combined average is greater in reading and spelling. The greater difference in results is shown in arithmetic calculation, in which 57.9 per cent of the experimental pupils exceed against 50.9 per cent control. It will be further noted that in arithmetic calculation the number of pupils in each grade who exceed the combined average gain is greater in the experimental than in the control.

When all subjects are combined and the results computed by grade, the number of pupils who exceed the combined average is greater in Grades 4 and 5 in the experimental group and in Grade 3 in the control group.

When all pupils are combined by grade and subject, the number of those exceeding the combined average gain is greater in the experimental than in the control.

Percentage of Pupils in Experimental and Control Groups Who Gain

The number of pupils who made gains in the weighted composite results by grade and subject was noted and the percentage this number was of the total experimental group and of the total control group was computed. The results are shown in Table XXIX.

When computed by grades in each subject, the experimental groups exceed in five instances, the control in three, while the results are the same in four instances.

When the totals are computed for each subject, the experimental groups exceed in arithmetic calculation and arithmetic problem solution, while the control groups exceed in reading and spelling.

When the totals are computed by grades regardless of subject, the experimental exceeds in Grade 4 while the control exceeds in Grades 3 and 5.

Summary.—The gains of the experimental groups exceed in a greater number of instances than do those of the control group when the gains are measured by single tests. When the pupils are grouped in individual classes, the experimental exceed 76 times and the control 43 times; when the

TABLE XXIX

PERCENT OF PUPILS IN EXPERIMENTAL AND CONTROL GROUPS WHO GAIN

Grade	Experimental			Control		
	Num. Gain	Total	%	Num. Gain	Total	%
			READING			
3	29	29	100	29	29	100
4	39	39	100	35	39	89.7
5	24	33	72.7	30	33	90.9
Total	93	101	90.9	94	101	93.1
			SPELLING			
3	33	35	94.2	34	35	92.9
4	45	51	88.2	45	51	88.2
5	38	43	88.3	41	43	95.6
Total	116	129	89.9	120	129	93.0
			ARITHMETIC CALCULATION			
3	27	27	100	26	27	96.3
4	36	40	90	36	40	90
5	35	37	94.6	35	37	94.6
Total	98	104	94.2	97	104	93.2
			ARITHMETIC PROBLEMS			
3	29	33	87.8	31	33	93.9
4	45	57	78.9	45	57	78.9
5	38	55	69.1	33	55	60.0
Total	113	145	77.9	109	145	75.1
			GRADE 3			
Reading	29	29	100	29	29	100
Spelling	33	35	94.2	34	35	92.9
Ar. Calc.	27	27	100	26	27	96.3
Ar. Prob.	29	33	87.8	31	33	93.8
Total	118	124	95.1	120	124	96.7
			GRADE 4			
Reading	39	39	100	35	39	89.7
Spelling	45	51	88.2	45	51	88.2
Ar. Calc.	36	40	90	36	40	90
Ar. Prob.	45	57	78.9	45	57	78.9
Total	165	187	88.2	161	187	86.1
			GRADE 5			
Reading	24	33	72.7	30	33	90.0
Spelling	38	43	88.3	41	43	95.6
Ar. Calc.	35	37	94.6	35	37	94.6
Ar. Prob.	38	55	69.1	33	55	60.0
Total	136	168	80.9	139	168	82.6

classes are grouped in grades, the experimental exceed 22 times and the control 8 times; when the classes are grouped according to type school, the experimental exceed 18 times and the control 12; when all pupils are grouped in one experimental and one control group, the experimental group exceeds 8 times and the control 2 times.

The gains of the experimental group exceed in a greater number of instances than do those of the control group when the gains are measured by composites (weighted and unweighted) of the tests in the four subjects—reading, spelling, arithmetic calculation, and arithmetic problem solution. When the classes are grouped into grades, the experimental exceed 14 times and the control exceed 10 times; when the classes are grouped according to type school, the experimental exceed 16 times and the control 8 times; when all pupils are grouped in one experimental and one control group, the experimental group exceeds 6 times and the control 2 times. It is thus seen that when the results are judged merely by the number of instances in which the experimental or the control group exceeds, the balance in each of the different groupings is in favor of the experimental group.

The results of the experimental group are slightly better in the 4th grade than in the 3rd and 5th. When measured by single tests, the experimental group exceeds 7 times in Grade 3, 8 times in Grade 4, and 7 times in Grade 5. When measured by weighted and unweighted composites, the experimental group exceeds 4 times in Grade 3, 6 times in Grade 4, and 4 times in Grade 5.

The differences in the results according to type school situations are very slight. When measured by single tests, the city groups exceed 5 times, the campus groups 6 times, and the rural groups 5 times. When measured by weighted and unweighted composite scores, the city groups exceed 5 times, the campus groups 6 times, and the rural groups 5 times. The order of success is, therefore, campus, city and rural, though the differences are too slight to make this of much significance.

The results are summarized in Table XXX.

TABLE XXX

NUMBER OF TIMES EXPERIMENTAL AND CONTROL GROUPS EXCEED IN AVERAGE GAIN WHEN RESULTS ARE MEASURED BY SINGLE TESTS AND BY WEIGHTED AND UNWEIGHTED COMPOSITES, PUPILS GROUPED BY GRADE, TYPE SCHOOLS, AND IN TOTAL EXPERIMENTAL AND CONTROL GROUPS

Subject	Grade 3		Grade 4		Grade 5	
	*Ex. Ex.	C. Ex.	Ex. E.	C. Ex.	Ex. Ex.	C. Ex.
AVERAGE GAINS SINGLE TESTS						
Reading	8	1	7	2	3	0
Spelling	3	3	3	3	1	1
Ar. Calc.	9	0	5	4	3	0
Ar. Prob.	2	3	2	4	1	1
AVERAGE GAINS WEIGHTED AND UNWEIGHTED COMPOSITES						
Reading	2	4	4	2	2	0
Spelling	2	4	4	2	0	2
Ar. Calc.	6	0	4	2	2	0
Ar. Prob.	4	2	4	2	2	0

* Ex. Ex.—Experimental exceeds.
C. Ex.—Control exceeds.

Table XXX indicates that the most consistent gain, measured both by single test and composite test scores were made in arithmetic calculation, while Table XXVI indicates that these gains were likewise the most reliable. The least satisfactory results made by the experimental group were found in spelling, the experimental and control groups being equal in the number of excess gains made in single tests and the balance being in favor of the control in the composite results. In reading, the balance is strongly in favor of the experimental group in the single test results and fairly so in the composite results. In arithmetic problem solution this last condition is reversed, the balance being in favor of the experimental group in the composite results, while in the single test results the balance is in favor of the control. It will be noted that in no case is the balance consistently in favor of the control group, but that in two cases, strongly so in the case of arithmetic calculation and fairly so in the case of reading, the balance is in favor of the experimental group.

When the results are studied in terms of the percentage of pupils in experimental and control groups who exceed the average weighted composite gain of the experimental and

control groups combined, the foregoing results are corroborated in so far as the experimental grades exceed the control in 7 as against 4 instances; in that the highest percentage of pupils exceeding in the experimental group is made in arithmetic calculation; and in the fact that the highest percentage exceeding in the experimental group when the pupils are grouped by grade is in Grade 4.

When the results are studied in terms of the percentage of pupils in the experimental and control groups who gain, the experimental exceed the control in the greater number of instances: Experimental exceed 5 times, control exceed 3 times, and results are identical in 4 instances. Again, the experimental group exceeds in arithmetic calculation and problem solution, while the control exceeds in reading and spelling. Again, also, the excess gain is made by the experimental group in Grade 4, the control group exceeding in Grades 3 and 5.

It would thus seem that in every measure used, the results favor the experimental group in arithmetic calculation in so far as subject is concerned and favor Grade 4 in so far as grade is concerned. In no other subject and in neither of the other grades are the results so consistent. In no case are consistent results found in favor of the control group. Furthermore, in general, excess gains are found in favor of the experimental groups. While the tables and figures of this chapter indicate that the differences in gains made by the experimental and control groups are not large, the balance on the whole favors the experimental group.

2. SUBORDINATE CRITERION: THE SUCCESS OF THE STUDENT TEACHER

The teaching success of the students was measured by the final student teaching grade. This grade is assigned by the director of student teaching and is based upon the reports made to him by the training teachers, various normal school instructors who supervise the students, and upon his own observation. The grade is expressed on a scale of 1 to 7, 1 being the lowest and 7 the highest grade assigned.

Distributions of the student teaching grades of the equated groups of students were made. The results were as follows:

	Experimental	Control
City Students		
Mean	5.06	5.03
S. D.	1.28	1.40
Cases	31	31
County Students		
Mean	5.45	4.95
S. D.	2.15	1.28
Cases	20	20
Total Students		
Mean	5.22	5.00
S. D.	1.09	1.35
Cases	51	51

The distributions for the above are shown in Figures 27, 28, and 29. While the differences are slight, in each case the difference is in favor of the experimental group.

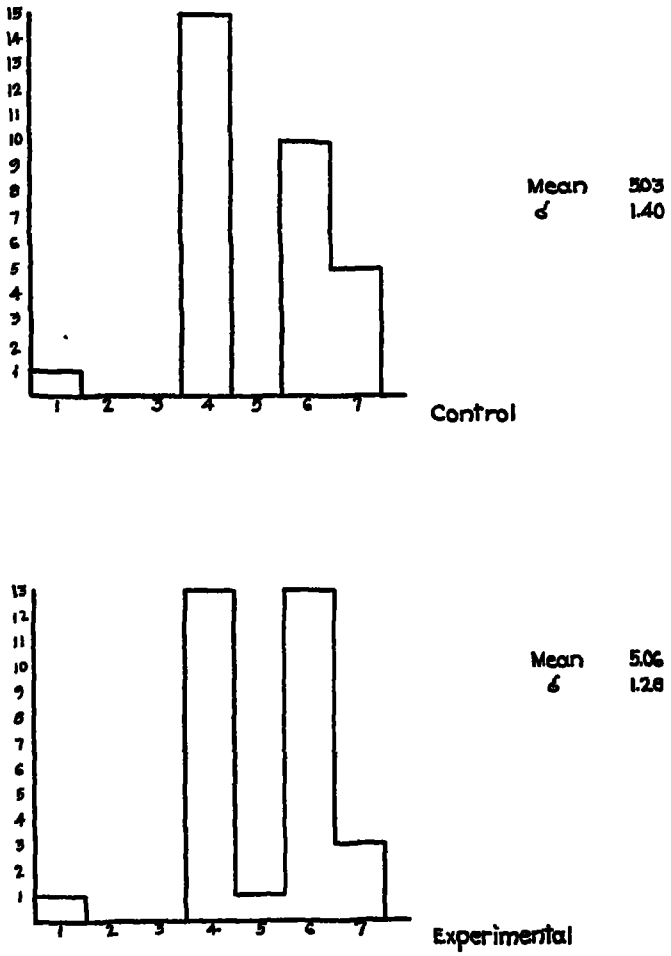

FIGURE 27.—Final student teacher rating of 31 experimental and 31 control city students.

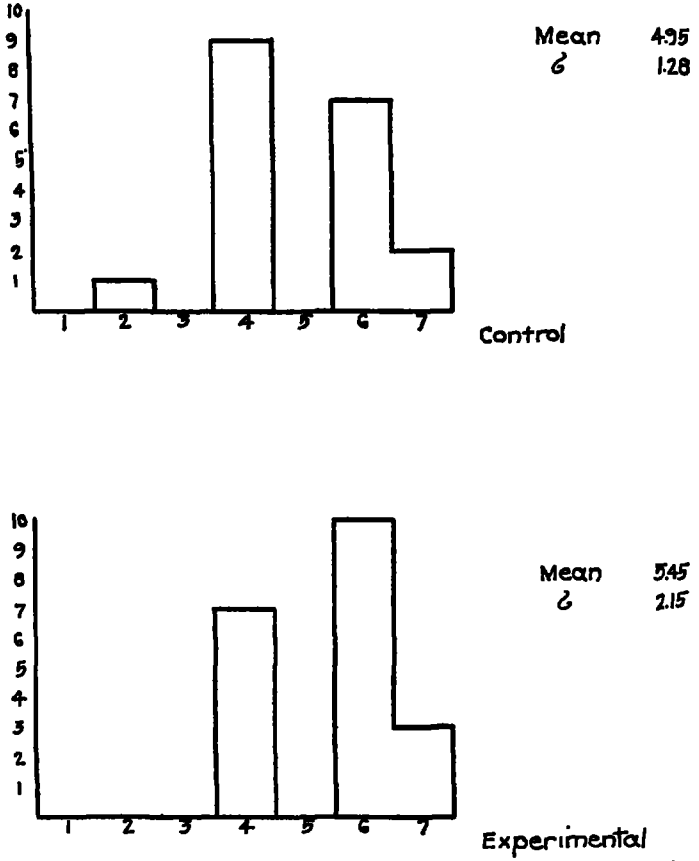

FIGURE 28.—Final student teacher rating of 20 experimental and 20 control rural students.

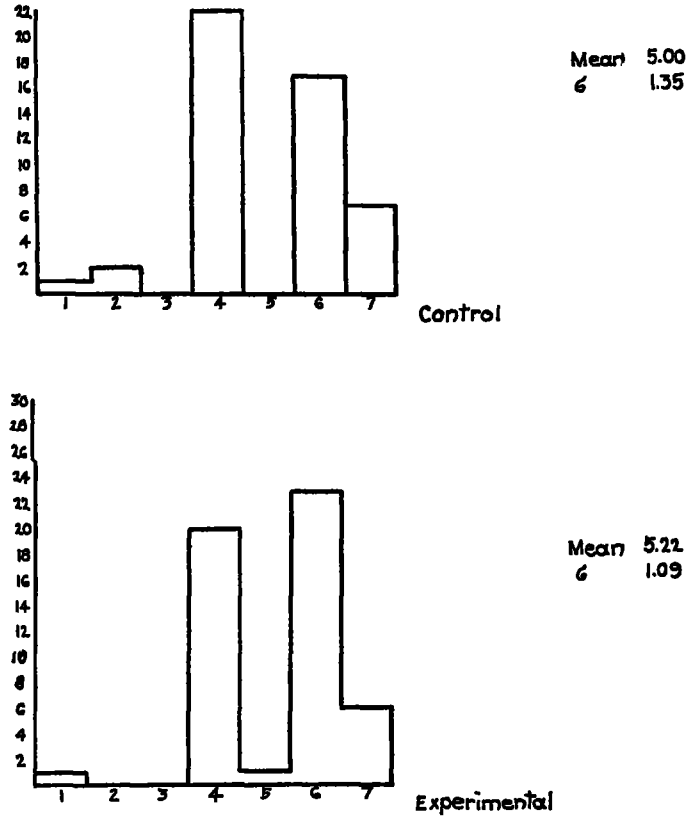

FIGURE 29.—Final teacher rating of all student teachers: 51 experimental and 51 control.

CHAPTER IX

The Measurement of Variables

I. IN RELATION TO GAINS MADE BY PUPILS

The Relation of Gains to Initial Status.—The correlation was computed by the product-moment formula between the initial status in the weighted composite scores in each subject and the gains of the experimental and control groups of each grade. The results (Table XXXI) indicate that on the

TABLE XXXI

CORRELATION BETWEEN INITIAL STATUS IN WEIGHTED COMPOSITE SCORES AND GAINS OF EXPERIMENTAL AND CONTROL GROUPS OF EACH GROUP

	Grade 3					Grade 4				
		E		C			E		C	
	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.
Reading	29	—.15	.1255	†—.50	.1248	39	—.21	.1079	—.29	.1072
Arith. Cal.	27	—.24	.1295	.34	.1305	40	—.27	.1066	—.22	.1066
Arith. Prob.	33	—.14	.1174	—.14	.1801	57	†—.44	.0890	†—.34	.0890
Spelling	35	.19	.1140	—.223	.1086	51	.102	.0938	—.07	.0944
Av. by Grades	124	—.085	.0603	—.131	.0542	187	†—.179	.0477	†—.131	.0484

	Grade 5					Average by Subjects				
		E		C			E		C	
	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.
Reading	33	—.23	.1174	—.25	.1166	101	†—.197	.0632	†—.35	..058
Arith. Cal.	37	.01	.1106	—.10	.1106	104	—.17	.0647	—.007	.0661
Arith. Prob.	55	.26	.0911	—.11	.0902	145	—.08	.055	†—.15	.1546
Spelling	43	.108	.1018	—.10	.1032	129	.048	.059	—.131	.058
Av. by Grades	168	.037	.0519	—.14	.0756					

* C.—Cases.
† —Reliable coefficients.

whole the coefficient is low and unreliable. As judged by the relative size of the coefficient and P. E. of the 40 coefficients calculated, but 8 can be considered reliable.

In the experimental group there are 7 positive and 12 negative coefficients, while in the control group there are 1 positive and 18 negative coefficients. In 14 instances the coefficients of the experimental group are higher, in 4 those of the control are higher, and in 1 instance the coefficient is the same. Furthermore, the average of the coefficients of the various subjects is higher in the experimental than in the control in each of the three grades and is higher for 3 of the 4 subjects when the pupils are massed in one experimental and one control group in each subject.

Reading	—Experimental higher 3 times; control, 0 time.
Arith. Calc.	—Experimental higher 1 time ; control, 2 times.
Arith. Prob.	—Experimental higher 1 time ; control, 1 time.
Spelling	—Experimental higher 3 times; control, 0 time.
Grade 3	—Experimental higher 2 times; control, 1 time.
Grade 4	—Experimental higher 2 times; control, 2 times.
Grade 5	—Experimental higher 4 times; control, 0 time.
Total by grades	—Experimental higher 3 times; control, 0 time.
Total by subject	—Experimental higher 3 times; control, 1 time.

In general, there is a greater degree of correlation between initial status and gains in the experimental than in the control groups. The experimental and control groups differ in the relative relationship between the gains and initial status of the 4 subjects tested and in the relative relationship between initial status and gains in the 3 grades. Not much importance can be attached to these results, however, because of the low reliability indicated.

In commenting upon the effect of errors of measurement on the correlation of initial status with gains, Godfrey H. Thomson says:

"As Thorndike showed in a recent article, if there is no real correlation between an initial value of some essential function and its gain over a certain period, then because of errors of measurement, there will appear to be a negative correlation between these quantities as measured. And for the same reason, any real correlation there may be, will be reduced in the measurement." [1]

Thomson proposes the following formula as a correction for errors of measurement when initial status is correlated with gains:

[1] Thomson, Godfrey H. "A Formula to Correct for the Effect of Errors of Measurement on the Correlation of Initial Values with Gains," Journal of Experimental Psychology, VII, No. 4, 1924.

$$r_{ag} = \frac{r_{xy} + \frac{\sigma_x}{\sigma_y}(1 - r_x)}{\frac{1}{\sigma_y}\sqrt{r_x\{\sigma_y^2 - \sigma_x^2(1 - r_x) - \sigma_z^2(1 - r_z)\}}}$$

in which:

a—true initial value
g—true gain
ag—true final value
rx—reliability coefficient of initial value
rz—reliability coefficient of final value
x—measured initial value
y—measured gain
z—measured final value

Since the formula calls for the use of reliability coefficients, effort was made to obtain all available reliability coefficients of the various tests used. The data on the reliability of these tests (except Stanford Achievement Test) are scanty. The coefficients are not the best one would desire, but are the best approximations that the writer could obtain by a reasonable expenditure of time. The reliability found for the composite scores is indicated in Table XXXII.

Using the reliability coefficients thus derived for values r_x and r_z and substituting the other values in Thomson's formula, the raw coefficients between Initial Status and Gains of Table XXXI were corrected and the coefficients of Table XXXIII were found.

The first difference to be noted between the raw and the corrected coefficients is a slight increase in reliability. While but 8 of the raw coefficients can be considered reliable, 14 of the corrected coefficients can be considered reliable.

The number of negative coefficients is reduced in both groups, but markedly so in the control group: Experimental, 14 positive and 5 negative coefficients; control, 14 positive and 5 negative coefficients.

In 13 instances, the coefficients of the experimental group are higher and in 6 instances the coefficients of the control group are higher. When the coefficients are averaged by grades, in 2 of the 3 grades the coefficients of the experi-

TABLE XXXII

RELIABILITY COEFFICIENTS OF THE VARIOUS TESTS AND THE WEIGHTED MEAN RELIABILITY COEFFICIENTS OF THE COMPOSITES

READING

Grade	Stanford	Thorndike McCall	Monroe	Weighted Mean
3	.93	.84	.73	.88
4	.93	.75	.64	.84
5	.91	.69	.55	.89

ARITHMETIC CALCULATION

Grade	Stanford	Monroe	Courtis	Weighted Mean
3	.91	.75	.75	.83
4	.91	.75	.75	.84
5	.87	.75	.75	.81

ARITHMETIC PROBLEM SOLUTION

Grade	Stanford	Buckingham	Weighted Mean
3	.86	.80	.84
4	.87	.80	.85
5	.88	.80	.85

SPELLING

Grade	Stanford	Morrison McCall	Weighted Mean
3	.89	.80	.86
4	.89	.80	.86
5	.91	.80	.87

INTELLIGENCE

Grade	Illinois	Trabue	Weighted Mean
3	.77	.81	.79
4	.79	.84	.82
5	.84	.84	.84

TABLE XXXIII

CORRELATION BETWEEN INITIAL STATUS IN WEIGHTED COMPOSITE SCORES AND GAINS OF EXPERIMENTAL AND CONTROL GROUPS OF EACH GRADE

Grade 3					Grade 4					Grade 5					Average by Subjects				
	E		*C	C	*C	E		*C	C	*C	E		*C	C	*C	E		*C	C
	.11	±.125	29	—.012	±.125	39	—.013	±.1079	39	—.18	±.1079	33	—.09	±.1174	33	—.05	±.1066	101	.007	±.0667	101	—.080	±.0661
	—.18	±.1295	27	†.53	±.0946	40	—.18	±.1066	40	—.05	±.1066	37	.29	±.1015	37	.09	±.1390	104	—.023	±.0661	104	†.190	±.0634
	.12	±.1174	33	.041	±.1174	57	†.29	±.0938	57	†.28	±.0938	55	†.48	±.1025	55	.15	±.1032	145	†.296	±.0505	145	†.157	±.0539
	†.36	±.1140	35	.078	±.1133	51	†.39	±.0219	51	.19	±.0938	43	†.691	±.0607	43	.087	±.1017	129	†.481	±.0951	129	.118	±.058
	.102	±.0506	124	†.159	±.0590	187	†.122	±.0485	187	.0602	±0506	168	†.342	±.0503	168	.069	±.0458						

*C.—Cases. † Reliable coefficients.

TABLE XXXIV

CORRELATION BETWEEN MENTAL AGE AND WEIGHTED COMPOSITE GAINS IN THE VARIOUS SUBJECTS BY GRADES

	Grade 3				Grade 4				Grade 5				Average by subjects			
	E		C		E		C		E		C		E		C	
	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.
	29	—.021	.1221	—.33	.1248	39	†.350	.0951	.083	.1072	33	—.099	.1165	.050	.1147	101	.076	.0667	—.066	.099
al.	27	—.148	.1221	.228	.1227	40	.124	.1052	.079	.1066	37	.114	.1093	.031	.1106	104	.030	.0661	.060	.065
ob.	33	—.045	.1147	.138	.1153	57	—.024	.0890	—.148	.0870	55	†.374	.0782	.214	.0863	145	.102	.064	.068	.055
	35	.099	.1133	.209	.1085	51	†326	.0843	†.329	.0836	43	†303	.0931	.192	.0992	129	†.242	.0559	†.243	.0559
	124	—.029	.0600	.061	.0600	187	†.194	.0472	.046	.0492	168	†.173	.0505	.121	.0512					

*C.—Cases. † Reliable coefficients.

mental group are higher; when averaged by subjects in 3 of the 4 subjects the experimental group is higher:

Reading	—Experimental higher 2 times; control, 1 time.
Arith. Calc.	—Experimental higher 1 time; control, 2 times.
Arith. Prob.	—Experimental higher 3 times; control, 0 time.
Spelling	—Experimental higher 3 times; control, 0 time.
Grade 3	—Experimental higher 3 times; control 1 time.
Grade 4	—Experimental higher 3 times; control, 0 time.
Grade 5	—Experimental higher 3 times; control, 1 time.
Total by grades	—Experimental higher 2 times; control, 1 time.
Total by subjects	—Experimental higher 3 times; control, 1 time.

Correcting the coefficients raised them on an average of .254 points in the experimental and .247 points in the control groups. The minimum increase in the experimental group was .06 points and the maximum increase was .58 points. The minimum increase in the control group was .11 points and the maximum, .62 points.

As in the case of the raw coefficients, there is a greater degree of correlation between initial status and gains in the experimental than in the control groups. Again, the experimental and control groups differ in the relative degree of correlation between the gains and initial status of the 4 subjects and in the relative degree of correlation between initial status and gains in the 3 grades. While there is some variation in the results of the raw and the corrected coefficients, in general the relative standing of the experimental and the control groups remains the same. It is interesting to note that the only instance in which the control group exceeds the experimental is in the case of arithmetic calculation, the subject in which the gains are most consistently in favor of the experimental group.

While the reliability of the corrected coefficients is higher than that of the raw coefficients, it must be remembered that but 14 of the 38 coefficients are large enough to indicate any positive or negative correlation. Of these 14 significant coefficients, however, the 9 significant coefficients of the experimental group exceed those of the control group, while of the 5 significant coefficients of the control group, but 3 exceed those of the experimental group.

The Relation of Gains to Mental Age

Coefficients of correlation were computed between weighted composite gains in each subject by grades and mental age, the mental age being the average of the Trabue Mentimeter and the Illinois Intelligence Examination. Data for correction of the raw coefficients were not available and, therefore, the results must be interpreted with the limitations of the uncorrected coefficients in mind. The results presented in Table XXXIV indicate generally low and unreliable coefficients, but 7 of the experimental and 2 of the control reaching a standard of reliability, while 6 of the experimental and 3 of the control are negative. The coefficients of the experimental groups exceed those of the control in 11 instances, while those of the control exceed the experimental in 8 instances. When the pupils are grouped by grades with all subjects combined, the coefficients of the experimental group are greater in Grades 4 and 5, while that of the control is greater in Grade 3. When all the grades are combined by subject, the coefficients of the experimental group are greater in reading and arithmetic problems, while those of the control are greater in arithmetic calculation and spelling.

Reading	—Experimental higher 2 times;	control, 1 time
Ar. Calc.	—Experimental higher 2 times;	control, 1 time
Ar. Prob.	—Experimental higher 2 times;	control, 1 time
Spelling	—Experimental higher 1 time;	control, 2 times
Grade 3	—Experimental higher 1 time;	control, 3 times
Grade 4	—Experimental higher 3 times;	control, 1 time
Grade 5	—Experimental higher 3 times;	control, 1 time
Total by grades	—Experimental higher 2 times;	control, 1 time
Total by subjects	—Experimental higher 2 times;	control, 2 times

The Relation of Gains to Mental Age with Chronological Age Constant

The relation between weighted gains and mental age with chronological age constant in the various subjects and grades was computed according to the formula for partial correlation:

$$r_{12.3} = \frac{r_{12} - r_{13.23}}{\sqrt{1 - r^2_{13}}\sqrt{1 - r^2_{23}}}$$

in which

r_{12} = coefficient of correlation between mental age and gains.
r_{13} = coefficient of correlation between mental age and chronological age.
r_{23} = coefficient of correlation between gains and chronological age.
$r_{12.3}$ = coefficient of correlation between mental age and gains with chronological age constant.

The correlation between mental age and chronological age is shown in Table XXXV. The results, in general, show low, negative, unreliable coefficients. In the experimental group there are 2 positive and 17 negative coefficients, while in the control group there are 6 positive and 13 negative coefficients. In the experimental group, 11 coefficients might be considered reliable, while in the control but 6 might be considered reliable. The significant thing to be noted is that in but one case the coefficient in the experimental group is higher than in the control. Since so many of the results are negative, the higher coefficients of the control group can be interpreted to indicate less disparity between mental and chronological ages than in the experimental group.

The correlation between chronological age and gains is shown in Table XXXVI. The coefficients are generally low and unreliable. In the experimental group there are 4 positive and 15 negative coefficients, while in the control group there are 1 positive and 18 negative coefficients. In 11 instances, the coefficients of the experimental group may be considered reliable and in 9 instances those of the control group may be considered reliable. In 12 instances, the coefficients of the experimental group exceed those of the control, and in 7 instances the coefficients of the control exceed those of the experimental. The coefficients of the experimental exceed the control in the entire group in reading and arithmetic problems, while those of the control exceed in arithmetic calculation and spelling. The coefficients of the experimental are greater in Grades 3 and 4, while those of the control are greater in Grade 5. It would seem that on the whole,

TABLE XXXV

CORRELATION BETWEEN MENTAL AGE AND CHRONOLOGICAL AGE OF EXPERIMENTAL AND CONTROL GROUPS IN EACH GRADE

	Grade 3					Grade 4					Grade 5					Average by subjects				
		E		C			E		C			E		C			E		C	
	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.	*C	r	P. E.	r	P. E.
Reading	29	.023	.0121	.025	.1248	39	.012	.1079	.121	.1060	33	—.227	.1113	—.123	.1153	101	—.192	.0640	—.175	.0647
Arith. Cal.	27	—.117	.1481	—.115	.1275	40	—.322	.0951	.185	.1032	37	—.431	.0904	†—.321	.0992	104	†—.290	.0600	—.083	.0654
Arith. Prob.	33	—.169	.1140	—.009	.1173	57	—.267	.0830	†.250	.0890	55	—.361	.0796	†—.323	.0809	145	†—.265	.0519	—.982	.0532
Spelling	35	—.273	.1059	.035	.1126	51	†—.398	.0796	—.004	.0850	43	—.232	.0978	†—.375	.0884	120	†—.301	.0539	—.114	.0586
Av. by Grades	124	†—.134	.0593	—.016	.0600	187	†—.325	.0438	†.138	.0480	168	†—.312	.069	†—.285	.071					

TABLE XXXVI

CORRELATION BETWEEN WEIGHTED GAINS IN THE VARIOUS SUBJECTS AND CHRONOLOGICAL AGE, PUPILS GROUPED BY GRADES.

	Grade 3					Grade 4					Grade 5					Average by subjects				
		E		C			E		C			E		C			E		C	
	*C	r	±P. E.	r	±P. E.	*C	r	±P. E.	r	±P. E.	*C	r	±P. E.	r	±P. E.	*C	r	±P. E.	r	±P. E.
Reading	29	†—.218	.0607	—.318	.1126	39	.148	.1059	†—.443	.0863	33	—.125	.1153	—.299	.1066	101	—.065	.0667	†—.350	.0586
Arith. Cal.	27	†—.450	.1032	†—.360	.1126	40	†—.374	.0917	—.233	.1005	31	.061	.2536	—.155	.1079	104	†—.254	.0613	†—.249	.0619
Arith. Prob.	33	†.345	.1032	—.178	.1133	57	.004	.0890	—.074	.0890	55	†—.395	.0769	.107	.0897	145	—.046	.0553	—.048	.0553
Spelling	35	—.012	.1086	—.037	.1141	51	†—.280	.0863	†—.305	.0863	43	†—.407	.0863	—.241	.0964	129	†—.233	.0559	†—.194	.0566
Av. by Grades	124	—.084	.0500	†—.223	.0573	187	†—.123	.0485	†—.267	.0451	167	†—.217	.0492	†—.172	.0505					

TABLE XXXVII

CORRELATION BETWEEN MENTAL AGE AND GAINS WITH CHRONOLOGICAL AGE CONSTANT

	Grade 3					Grade 4					Grade 5					Average by subjects				
		E		C			E		C			E		C			E		C	
	*C	r	±P. E.	r	±P. E.	*C	r	±P. E.	r	±P. E.	*C	r	±P. E.	r	±P. E.	*C	r	±P. E.	r	±P. E.
Reading	29	—.016	.1187	—.234	.1187	39	†.353	.0944	.153	.1052	33	—.132	.1155	.014	.1174	101	.068	.0668	†—.22	.0668
Arith. Cal.	27	—.229	.1228	†.636	.0769	40	.004	.1066	.053	.1059	37	.155	.1079	—.020	.1106	104	†—.23	.06610	†.223	.0627
Arith. Prob.	33	.014	.1174	.139	.1153	57	—.023	.0890	—.13	.0890	55	†.270	.0843	†.264	.0843	145	.087	.0553	.091	.0553
Spelling	35	.099	.1126	.212	.1086	51	†.243	.0890	†.364	.0816	43	.234	.0971	.113	.1018	129	†.192	.0567	†.229	.0573
Av. by Grades	124	—.033	.0600	.129	.0486	187	†.149	.0485	.110	.0513	168	.132	.0513	.092	.0519					

*C.—Cases. † Reliable coefficients.

the factor of age has had somewhat more influence on the results of the experimental than on those of the control group.

The correlation between mental age and gains with chronological age constant is shown in Table XXXVII. The results indicate generally low and unreliable coefficients. In the experimental group there are 13 positive and 6 negative coefficients, while in the control group there are 15 positive and 4 negative coefficients. Six of the experimental and 6 of the control coefficients can be considered reliable. The coefficient of the experimental group is greater in 9 cases and that of the control group is greater in 10 cases. When grouped regardless of subject according to grade, the coefficient of the experimental group is greater in Grades 4 and 5 and that of the control in Grade 3. When grouped by subject, the coefficient of the experimental is greater in reading, while the coefficient of the control is greater in arithmetic calculation, arithmetic problem solution, and spelling.

Reading	—Experimental higher 2 times;	control, 1 time
Ar. Calc.	—Experimental higher 1 time;	control, 2 times
Ar. Prob.	—Experimental higher 2 times;	control, 1 time
Spelling	—Experimental higher 1 time;	control, 2 times
Grade 3	—Experimental higher 1 time;	control, 3 times
Grade 4	—Experimental higher 2 times;	control, 2 times
Grade 5	—Experimental higher 3 times;	control, 1 time
Total by grades	—Experimental higher 1 time;	control, 2 times
Total by subjects	—Experimental higher 1 time;	control, 3 times

When the influence of chronological age is removed, the factor of mental age bears a somewhat greater relationship to gains in the control than in the experimental group. It has been shown (1) that the correlation between mental age and gain is greater in the experimental group; (2) that the correlation between chronological age and gain is negative but nearer zero in the experimental; (3) that the correlation between mental age and chronological age is negative but nearer zero in the control; (4) that the correlation between mental age and gain with chronological age constant is slightly greater in the control.

The Relation of Gains to Amount of Time Given to Test-Determined Instruction

Correlations by the rank method (Spearman's Rank Difference) were computed between the gains made in each test and the amount of time given to test-determined instruction in each subject by the various experimental classes. The results are shown in Table XXXVIII.

TABLE XXXVIII

Correlation between Time and Gains Made in Separate Tests by Individual Experimental Classes

Time and	No. Classes	r
Reading:		
Stanford	12	.124
Thorndike McCall	12	—.158
Monroe	12	—.541
Average		—.191
Arithmetic Calculation:		
Stanford	12	.158
Monroe	12	.158
Courtis	12	—.018
Average		.099
Arithmetic Problem Solution:		
Stanford	12	—.071
Buckingham	12	—.354
Average		—.213
Spelling:		
Stanford	12	.141
Morrison McCall	12	.642
Average		.381

Table XXXVIII indicates that there is little relation between the amount of time spent on test-determined instruction by the various experimental classes and the gains made in separate tests. The small number of cases—12 classes—gives little significance to these results. The order of relationship between time and gain from high to low is as follows:

1. Spelling —Av. r .381
2. Ar. Calc. —Av. r .099
3. Reading —Av. r —.191
4. Ar. Prob.—Av. r —.213

The relation between time spent in test-determined instruction and gains made in weighted composite scores when

pupils are grouped according to grade and type school situations is shown in Table XXXIX. This table was derived from Tables XIV and XV (Chapter VII) and Tables XXVI and XXVII (Chapter VIII). From Tables XIV and XV were obtained the rank in amount of time expended on experimental instruction by the various type schools and the various grades respectively. From Tables XVI and XVII were obtained the rank in order of gain made by the experimental groups as compared with the control groups by the various type school situations and by grades respectively.

TABLE XXXIX

COMPARISON OF TIME EXPENDED ON TEST-DETERMINED INSTRUCTION WITH GAINS MADE IN COMPOSITE WEIGHTED SCORES WHEN PUPILS ARE GROUPED ACCORDING TO GRADE AND TYPE SCHOOL SITUATION

	Reading		Spelling		Arith. Calc.		Arith Prob	
Grade	Gain	Time	Gain	Time	Gain	Time	Gain	Time
3	3	3	1	1	2	3	3	3
4	1	1	2	2	3	1	2	2
5	2	2	3	3	1	2	1	1
Type School								
City	2	3	3	1	1	2	3	3
Campus	3	1	2	2	2	1	2	1
Rural	1	2	1	3	3	3	1	2

N B—All figures in table indicate rank

Table XXXIX shows perfect agreement in rank in comparative gain in reading, spelling, and arithmetic problem solution with amount of time spent in test-determined instruction when the pupils are grouped according to grade, but no such agreement in the case of arithmetic calculation. No agreement is indicated between time and gain when the pupils are grouped according to type school situation.

Table XXVI (Chapter VIII) indicates that when all the pupils are grouped in one experimental and one control group excess gains over the control group are made by the experimental group in each subject except spelling, the rank being: 1. Arithmetic Calculation; 2. Reading; 3. Arithmetic Problem Solution; 4. Spelling.

Table XII and Table XV indicate that the relative amount of time directed to be spent on each of the four subjects and the relative average time actually spent on these subjects is identical, the rank in time for both being: 1. Reading; 2. Spelling; 3. Arithmetic Calculation; 4. Arithmetic Problem Solution. No relation, therefore, is indicated between the average amount of time spent on the various subjects by the experimental group as a whole and the results obtained. From all of the foregoing, it is evident that little relation between time and gain is shown.

The Relation of Intelligence, Scholarship, and Chronological Age of the Student Teacher to Gains Made by the Pupils

The averages of the I. E. R. Intelligence scores of the junior scholarship grades and of the chronological ages were computed for all the students who taught in the experimental and the control classes, the averages being indicated according to grade and type school situation. The results are shown in Table XL. Since the junior scholarship grades are expressed on a different scale for the city and rural students it was necessary to indicate them separately.

It was shown in Chapter VIII (pp. 87-93) that the school grades of the experimental group ranked in the degree to which the best weighted results were obtained as follows: 1. Grade 4; 2. Grade 3; 3. Grade 5. Table XL shows very little difference in average intelligence, scholarship, or chronological age, except in the case of chronological age which is somewhat higher in Grade 3 than in Grades 4 and 5. The rank within the slight differences shown is not in any case identical with the rank of the obtained results, the rank of the groups of students assigned to each grade being as follows:

	Scholarship		
Intelligence	City	Rural	Chronological Age
1. Grade 5	1. Grade 4	1. Grade 3	1. Grade 3
2. Grade 4	2. Grade 5	2. Grade 5	2. Grade 4
3. Grade 3	3. Grade 3	3. Grade 4	3. Grade 5

TABLE XL

AVERAGE INTELLIGENCE, SCHOLARSHIP, AND CHRONOLOGICAL AGE OF EXPERIMENTAL AND CONTROL GROUPS OF STUDENTS

	Experimental						Control					
	Intell.		Sch.		Chr. Age		Intell.		Sch.		Chr. Age	
Grade	N.	Av.	N.	Av.	N.	Av.	N.	Av.	N.	Av.	N.	Av.
3	44	283.75	*R. 23 C. 16	4.45 85.12	43	231.65	44	282.02	*R. 30 C. 14	4.56 84.43	48	225.03
4	69	285.84	R. 43 C. 18	4.43 85.88	68	227.89	67	288.97	R. 49 C. 18	4.53 85.50	69	223.58
5	68	286.57	R. 42 C. 17	4.44 85.35	67	226.55	66	296.16	R. 51 C. 16	4.64 85.68	68	220.57
City	51	331.06	51	85.46	49	237.89	47	343.29	48	87.06	47	237.14
Campus	36	269.86	30	4.47	36	227.97	33	265.72	30	4.42	33	222.20
Rural	39	266.89	33	5.06	38	233.39	39	273.38	40	4.68	41	221.09

* In this table R. indicates the average scholarship of all student teachers in the Campus and Rural schools, and C. of those in the City schools.

As measured by the difference between the experimental and control groups of students in each of these factors, the rank is as follows:

	Scholarship		
Intelligence	City	Rural	Chronological Age
1. Grade 5	1. Grade 3	1. Grade 4	1. Grade 3
2. Grade 4	2. Grade 4	2. Grade 3	2. Grade 5
3. Grade 3	3. Grade 5	3. Grade 5	3. Grade 4

Again, there is practically no correspondence between pupil results and comparative student status.

It was shown in Chapter VIII (pp. 87-93) that when the pupils were grouped according to type schools, the best results were obtained in the Campus School, the city and rural groups being identical in the number of instances in which the experimental groups exceeded the control. Table XL shows no corresponding superiority in the factors measured of the students who taught in this situation. The results on the whole indicate no relationship between the success of the pupils and the intelligence, scholarship, and chronological age of the student teacher.

Summary of Study of Variables in Relation to Gains Made by Pupils

1. There is a greater degree of correlation between initial status and gains in the experimental than in the control group, both when measured by raw and by corrected coefficients.
2. There is a greater degree of correlation between mental age and gains in the experimental than in the control group, but when chonological age is made constant, the relationship is slightly greater in the control than in the experimental.
3. Except when the pupils are grouped according to grade, but little relation is shown between gains and amount of time spent in test-determined instruction.
4. No relation is shown between gains made by pupils and the intelligence, scholarship, and chonological age of the student teachers.

2. THE MEASUREMENT OF VARIABLES IN RELATION TO THE FINAL RATING OF THE STUDENT TEACHERS

Correlation was computed between the final student teacher rating of the experimental and control groups of students and their I. E. R. Intelligence score, their junior scholarship grades, and their chronological age. Because of the difference

in the scale used, the junior scholarship grades had to be treated separately for the city and the rural students. The results are shown in Table XLI.

TABLE XLI

CORRELATION BETWEEN FINAL STUDENT TEACHER RATING AND CERTAIN OTHER STUDENT FACTORS

Final Student Teacher Rating	Experimental			Control		
	Cases	r	P. E.	Cases	r	P. E.
I. E. R. Intelligence	51	.038	.0944	51	.035	.0943
Jr. Schol. Grades City	30	.025	.1234	31	.242	.1139
Jr. Schol. Grades Rural	16	.518	.1233	20	.421	.1241
Chronological Age	51	.289	.0866	51	.144	.0924

Table XLI indicates:

1. That there is greater correlation in the experimental group between final teacher rating and intelligence, rural scholarship grades, and chronological age than in the control group, while the correlation in the control group is greater between final teacher rating and junior city scholarship grades.

2. That the highest correlations obtained, both in the experimental and the control groups are between final teacher rating and Junior Scholarship Grades Rural.

CHAPTER X

The Interpretation of the Data and Conclusions

I. THE MAIN CRITERION: THE PROGRESS OF THE PUPILS

The Results in General.—The problem of the present study was stated in the beginning as: *The place and value of training student teachers to use skilfully the results from general and diagnostic tests in the instruction of children.* The results that have been presented show, in general, a balance in favor of the experimental group; that is, that when the results are considered in terms of the number of times the experimental or the control group showed greater average gain without consideration of the subject or test in which the gains were made, the pupils taught by students trained to use the results of general and diagnostic tests showed greater gains when tested than the pupils taught by students not so trained. In order to bring the results together for convenience in interpreting them, the following summary is offered:

1. Measured by separate tests, greater gains are made by:
   Experimental—76; control—43 (classes)
   Experimental—15; control— 9 (type school)
   Experimental—20; control— 4 (grade)
   Experimental— 7; control— 1 (total)
   (Tables XVIa, b; XVII)

2. Measured by composite scores, weighted and unweighted, greater gains are made by:
   Experimental—14; control—10 (grades)
   Experimental—16; control— 8 (type school)
   Experimental— 6; control— 2 (total)
   (Tables XXVI; XXVII)

3. Measured by the number of pupils exceeding the average gain of the combined experimental and control groups, the results show:
   Experimental— 7; control— 3; same—1 (grades)
   Experimental— 2; control— 1 (all grades)
   Experimental— 2; control— 2 (all subjects)
   Experimental— 1; control— 0 (total)
   (Table XXVIII)

4. Judged by the number of instances in which the number of pupils who gain is greater in the experimental or in the control:
   Experimental— 4; control— 3; same—5 (grades)
   Experimental— 2; control— 2 (total)

In the first, second, and third measures given above, the balance is clearly in favor of the experimental group. In the fourth measure—the percentage of pupils who gain—the results are about the same in the experimental and control groups. Since the percentage of pupils who gained implies, conversely, that certain pupils showed no improvement or, as happened in by far the greater number of cases, that some pupils in each group made lower scores in the final than in the initial test, the results here may well be attributed, in part at least, to errors of measurement, enhanced, in the present study, by the small number of cases. The fourth item is, therefore, not considered of sufficient significance to outweigh the results of the other three measures.

As has been shown in the detailed study of results in Chapter VIII, the difference between the gains of the experimental and the control groups is slight. The smallness of the differences is attributable:

1. To the shortness of the period of test-determined instruction—19 weeks—this period subject to the many interruptions inevitable in the complicated program of the training centers.
2. To the fact that the results of this short period of instruction were measured not by specific tests of the functions exercised but by general standard achievement tests.

Under the above conditions, at best but a trend in one direction or the other could be expected. Such trend in favor of the experimental group is indicated in the above results. It is, therefore, concluded *that the number of instances in which the gains of the experimental group exceed those of the control is indicative of some value in training student teachers to use the results of general and diagnostic tests in their instruction of children.*

The influence of the variables of Initial Status and of Mental Age upon the results is summarized as follows:

1. The correlation between initial status and weighted composite gains is low but is higher in the experimental or in the control as follows:
   a. Uncorrected:
      Experimental higher—3; control higher—0 (grades)
      Experimental higher—3; control higher—1 (subjects)

b. Corrected:
Experimental higher—2; control higher—0 (grades)
Experimental higher—3; control higher—1 (subjects)
(Tables XXXI and XXXIII)

2. The correlation between mental age and weighted composite gains is low but shows higher coefficients in the experimental or the control as follows:
Experimental higher—2; control higher—1 (grades)
Experimental higher—2; control higher—2 (subjects)
(Table XXIV)

3. The correlation between mental age and weighted composite gains when chronological age is constant is low but shows higher coefficients in the experimental or in the control as follows:
Experimental higher—1; control higher—2 (grades)
Experimental higher—1; control higher—3 (subjects)

The results show a slightly higher degree of correlation between initial status and gains in the experimental than in the control groups. The same thing is true, but to a less degree, of the correlations between mental age and gains in the experimental than in the control groups. A very slightly higher degree of correlation between mental age and gains with chronological age constant in the control than in the experimental is shown.

While these results indicate in a very slight degree that the progress made by the pupils of the experimental group was more in accord with their mental age and their initial status in the subjects tested than by those in the control, the coefficients are so low and the differences are so small that no conclusion is justifiable.

The Results According to Subject.—The following summary of the relative gains made according to subject is offered:

1. When measured by separate tests, greater gains are made as follows:
a. Arithmetic Calculation:
Experimental—24; control—12 (classes)
Experimental— 9; control— 0 (grades)
Experimental— 5; control— 4 (type school)
Experimental— 3; control— 0 (total)
sum—41; sum—16
b. Reading:
Experimental—27; control— 9 (classes)
Experimental— 8; control— 1 (grades)
Experimental— 7; control— 2 (type school)
Experimental— 3; control— 0 (total)
sum—45; sum—12

c. Arithmetic Problems:
Experimental—11; control—12; same—1 (classes)
Experimental— 1; control— 5 (grades)
Experimental— 2; control— 4 (type school)
Experimental— 1; control— 1 (total)
sum—15; sum—22

d. Spelling:
Experimental—14; control—10 (classes)
Experimental— 3; control— 3 (grades)
Experimental— 3; control— 3 (type school)
Experimental— 1; control— 1 (total)
sum—21; sum—17
(Tables XVIa, b; XVII)

2. When measured by composite scores weighted and unweighted, greater gains are made by:

a. Arithmetic Calculation:
Experimental— 6; control— 0 (grades)
Experimental— 4; control— 2 (type school)
Experimental— 2; control— 0 (total)
sum—12; sum— 2

b. Reading:
Experimental— 2; control— 4 (grades)
Experimental— 4; control— 2 (type school)
Experimental— 2; control— 0 (total)
sum— 8; sum— 6

c. Arithmetic Problems:
Experimental— 4; control— 2 (grades)
Experimental— 4; control— 2 (type school)
Experimental— 2; control— 0 (total)
sum—10; sum— 4

d. Spelling:
Experimental— 2; control— 4 (grades)
Experimental— 4; control— 2 (type school)
Experimental— 0; control— 2 (total)
sum— 6; sum— 8
(Tables XXVI, XXVII)

Ranking the above in order of excess gain, we have:

1. By single tests:
1. Reading
2. Arithmetic Calculation
3. Spelling
4. Arithmetic Problems

2. By composites:
1. Arithmetic Calculation
2. Arithmetic Problems
3. Reading
4. Spelling

3. When measured by number of pupils exceeding average gain of combined experimental and control groups, the results are ranked as follows:

1. Arith. Calc. Exper., 57.9% exceed; control, 50.9%
2. Arith. Prob. Exper., 27.5% exceed; control, 26.2%
3. Reading Exper., 45.5% exceed; control, 46.5%
4. Spelling Exper., 42.5% exceed; control, 44.9%
(Tables XXVIIIa and b)

4. Measured by the number of instances in which the number of pupils who gain is greater in the experimental or control, the rank is as follows:

1. Arith. Calc.	Exper., 94.2%; control, 93.2%
2. Arith. Prob.	Exper., 77.9%; control, 75.1%
3. Reading	Exper., 90.9%; control, 93.1%
4. Spelling	Exper., 89.9%; control, 93.0%

(Table XXIX)

The rank by single test results is not in agreement with the rank by composites nor with the rank of the third and fourth measures above. Assuming the greater reliability of the composite results and in view of the fact that the rank in the third and fourth measures is in agreement with the rank of the composites, the latter can be assumed to represent the relative success of the experimental group. In each of the four main measures—gains by single tests, gains by composites, percentage exceeding average of combined experimental and control groups, percentage of pupils who gained—the most highly consistent gains are made by the experimental group in arithmetic calculation. A trend in favor of the experimental group is shown in reading and arithmetic problem solution, but not nearly to so great an extent as in arithmetic calculation. In the case of spelling, the results favor the control group.

The various subjects rank in order of degree of relationship between initial status and gains as follows:

1. Arith. Prob.	—Experimental higher, 3; control, 0
2. Spelling	—Experimental higher, 3; control, 0
3. Reading	—Experimental higher, 2; control, 1
4. Arith. Calc.	—Experimental higher, 1; control, 2

The results here are not in agreement with the results as measured by relative gains, since when measured by excess gains, the best results in the experimental group were in arithmetic calculation, while in the present instance, arithmetic calculation is the only subject in which the relationship between initial status and gains is greater in the control than in the experimental group. The problem is complicated by the relative difficulty of making a high score higher in these subjects as well as the particular stage of development in each when it is most difficult to make a high score higher. For instance, it may be more difficult in the fifth grade to make

a high score in arithmetic higher than it is to make a high score in reading higher. In the absence of definite knowledge as to the relative improvability of these subjects in the various school grades, and in view of the decided probability that such grade differences in improvability exist, the above results need not modify the conclusion indicated by the relative success of the experimental group in the different subjects, namely: *That when measured in terms of relative gains, both in individual test scores and in composite scores of the various subjects, the results indicate clearly that there is some value in teaching students to use the results of tests in arithmetic in their instruction of children, that there is a suggestion of such value in the case of reading and of arithmetic problem solution, but that no such value is indicated by the results in spelling; that in considering this conclusion, it must be kept in mind that the period of test-determined instruction was but 19 weeks in length and that no conclusion is offered as to what might have been the results had the experiment been carried over a longer period of time.*

In so far as it is possible to do so with the data at hand, attempt will be made to interpret the relative degree of success in the four subjects. The differences in results may be due to any one or more of a number of causes some of which are analyzable and some of which are not. Some of the possible causes of the differences in results, analysis of which will be attempted, are:

1. The relative amount of time spent on the various subjects.
2. The relative improvability of the subjects.
3. The relative worth of the test-determined instruction used.
4. The relative difficulty of the instruction to the students.
5. The relative degree to which the tests were measures of the instruction given.

Spelling.—At first sight it would seem that spelling is a subject which because of its definiteness and comparative simplicity of method should bring better results in the hands of the unskilled student teacher than any other. Instead, with an expenditure of next to the highest amount of time, the results are the poorest obtained. The bulletins (Chapter VI) indicate that the test-determined instruction in spelling had the following features:

1. Rousing desire for improvement through records of achievement, frequent testing, and establishment of definite goals.
2. Basing the work on an analysis of the difficulties shown in both the initial standard tests and the informal tests given weekly.
3. Grouping words around types of specific difficulties and varying the teaching emphasis according to the difficulty.
4. Supplementing the work in specific difficulties with words drawn from lists based upon common usage either by children or adults, such as the list compiled by Homer J. Smith and the "100 demons."
5. Keeping up interest through variety in drills used.
6. Grouping children according to needs shown.
(Bulletins 2, 5, 6, 10, 11, 13, 14, 17, 18.)

The above points conform, in general, to the principles of method in teaching spelling outlined by Ernest J. Horn.[1] Greater emphasis was placed, however, in the present study upon the grouping of words around type difficulties than Horn would seem to advocate in his statement: "Grouping by type of errors seems to have some merit in teaching words selected because of their similarity in type of error, but there is a tendency to include in such groupings words which need not be learned at all."[2] Gates likewise expresses doubt of the value of the method in the following statement: "Groupings by similarities of error, especially in review, are sometimes effective, but seldom worth the trouble."[3]

In support of the method of grouping about type difficulties, Tidyman says, "Grouping according to a common difficulty is a principle which appears to meet our needs. It is fundamentally pedagogical in character and broad enough to include all classes of words."[4] Again, Tidyman cites the following: "The experimental determination of the value of this plan of grouping was made by Dr. Wagner of the University of Pennsylvania who compared the growth in ability of two sixth-grade classes. In one class, the words were classified according to a common error and special attention was called to the words of the group. In the other class, the

[1] Horn, Ernest J. "Principles of Method in Teaching Spelling as Derived from Scientific Investigation," 18th Yearbook, National Society for the Study of Education, Part II, pp. 52-71.
[2] Horn, *Ibid.*, Part II, pp. 52-71.
[3] Gates, A. L. The Psychology of Reading and Spelling with Special Reference to Disability, Columbia University, 1922, p. 82.
[4] Tidyman, W. F. The Teaching of Spelling, 1919, p. 20.

words were presented in a chance order. In ten learning lessons the class following the grouping method showed a greater gain, amounting to about 20 per cent or 1.1 times the improvement of the ordinary class as calculated by the author. The conclusion is: 'Grouping words into lessons according to spelling difficulty is better than the ordinary plan of spelling lessons. It secures better daily lessons, better final results, and greater steadiness or constancy of correct spelling.'"[1]

Since the emphasis upon grouping by difficulty determined to a large extent the content of the test-determined instruction in spelling, the words specified in the bulletins to be taught were checked against the final tests given in order to determine the extent to which the tests actually tested the words taught. It was found that of the 180 words outlined to be taught in Grade 3, only 1 word appeared in the Morrison McCall Spelling Test, Form 4, and but 4 words appeared in the Stanford Achievement Test, Primary Examination, Form B, as follows:

Morrison McCall:	began
Stanford:	boat
	play
	miss
	read

Of the 180 words outlined for Grades 4 and 5, but 3 appeared in the Morrison McCall Spelling Test, Form 4, and but 4 appeared in the Stanford Achieement Test, Adanced Examination, Form B, Test 9, as follows:

Morrison McCall:	began
	belong
	over
Stanford:	boat
	child
	miss
	play
	read

It is evident that the tests did not measure ability to spell the words taught in the test-determined instruction. The degree to which the failure of the tests to measure the worth of

[1] Tidyman, *Ibid.*, p. 16.

the instruction is accountable for the comparatively poor spelling results is dependent upon the extent to which ability to spell certain words helps in the ability to spell other words. In this connection, Leta S. Hollingworth says: "English spelling is relatively resistant to learning because of the specific character of the connections to be made. Very few generalizations are possible, each word being to so great an extent a special matter." [1]

If English spelling connections are so highly specialized as Dr. Hollingworth indicates and as common observation corroborates, emphasis upon the selection of words in common use seems the reasonable basis for determining the content of a spelling course. As Courtis has said, "The basic aim of the spelling work is then, *not* the teaching of the spelling of the words of the course, but the development of the ideal of not making mistakes in spelling. . . . . The school must make sure that children have perfect control over spelling the vital words of the English language (the thousand most common words) and *they must arouse in the children such a desire to spell correctly that they will of themselves purpose to learn to spell all the other words they may need.*" [2]

While no record of the method used in the control classes was kept and while the methods used in the different situations varied, in general, the Horn-Ashbaugh method was used, supplemented by the teaching of other words as needed. In neither the experimental nor in the control group was more than ordinary emphasis placed upon the visual perception of the word, the importance of which is indicated by Gates [3] in the correlation of .50 between ability to discriminate between small differences between pairs of words and ability to spell. In commenting upon this, Homer B. Reed says, "The close dependence of spelling upon perception suggests that the most important condition for learning to spell is an accurate and

[1] Hollingworth, L. S. Special Talents and Defects, 1923, pp. 125, 126.
[2] Courtis, S. A. Teaching Spelling by Plays and Games, pp. 4, 5.
[3] Gates, Arthur I. The Psychology of Reading and Spelling, 1922, p. 31.

strong impression of the letters of the word. This may be acquired by adequate repetition and vivid presentation. Therefore, any method or device of teaching spelling that would increase either the frequency or the vividness of the presentation should help in improving ability to spell."[1]

The main point of difference between the method used in the experimental group and that of the control was the emphasis placed in the former upon grouping according to type difficulties. The results obtained are in agreement with the position of Gates and of Horn; i. e., that grouping spelling words according to type difficulty is of doubtful worth. The basis of selection of the spelling words unduly limited the character of the words taught and, since the ability to spell one word seems to have little power to affect the ability to spell other words and since, further, so few of the words taught actually occurred in the final tests given, the failure of the experimental group to make superior gains can be explained primarily upon the basis of a too-limited content. To make the conclusion sufficiently sound it would be necessary, of course, to have a check of the words actually taught in the control group against the final test given. Since this cannot be done, the conclusion could rest upon the assumption that a richer content of the course in the control group gave greater probability that more of the words taught appeared in the final tests.

The above interpretation and conclusion suggests that experimentation in how to teach student teachers to teach spelling be conducted, such experimentation to embody the following points of emphasis:

1. The analysis of difficulties disclosed by standard tests and frequent informal tests, such analysis to be followed by the correction of the errors made rather than by the teaching of other words of similar difficulty as was done in the present study.
2. Familiarity with the most scientifically-determined lists of words in common use and with the most scientifically-determined methods for teaching such words.
3. Definite instruction in how to teach children to want to spell correctly every word used and how to give them the necessary skill to learn such words without help from others.

[1] Reed, Homer B. Psychology of Elementary School Subjects, p. 230.

Arithmetic Calculation.—The work in arithmetic calculation, as outlined in the bulletins, was based upon:

1. Diagnosis according to the Courtis Diagnostic Test.
2. Grouping of children upon the basis of such diagnosis.
3. Rousing desire for improvement through records of achievement, frequent testing, and establishment of definite goals.
4. Frequent informal tests.
5. Study of children to note individual poor study habits and the correction of such habits.
6. Keeping up interest through variety of drill.

(Bulletins 3, 6, 7, 10, 11, 12, 13, 14, 15, 17, 19.)

Like the work in spelling, the test-determined instruction in arithmetic calculation was based upon the analysis of errors made. This analysis differed, however, from that of the analysis of the spelling, inasmuch as the arithmetic analysis was based upon a definite diagnostic test, The Courtis Cleveland Survey Test, designed to indicate weaknesses which must be overcome if skill in arithmetic is to be acquired.

While learning to spell one word is of doubtful value in learning to spell other words, skill in arithmetic calculation, on the other hand, is definitely dependent upon the mastery of elements which are used again and again in the more complex processes. Hence, the basic drill in fundamentals outlined in the bulletins can be expected to function in different arithmetic situations to a far greater extent than can the work in spelling be expected to function in other spelling situations. While it is impossible to check the test-determined instruction in arithmetic against the final tests given as was done in the case of spelling, such tests of necessity contained many of the identical arithmetic elements practiced.

It is, therefore, concluded that the results obtained are such as might be expected because of the hierarchical character of the elements of arithmetic, its improvability through definite practice, and the validity of the final tests as measures of the test-determined instruction. It is further concluded that the results justify the continuance of the training of student teachers to use the results of standard tests in arithmetic in their instruction of children.

Reading.—The test-determined instruction in reading as outlined in the bulletins included the following:

1. Rousing desire for improvement through records of achievement, frequent testing, and establishment of definite goals.
2. Grouping the children as need was shown by the results of the Stanford Achievement Test for special work in vocabulary, sentence meaning, and paragraph meaning.
3. Further analysis of reading difficulty through the use of the Jones Vocabulary Test, The Gray Oral Reading Test, and through observing evidences of poor reading habits.
4. The use (in the third term) of intensive reading exercises graded in difficulty on the basis of the reading level attained as indicated in the Thorndike McCall Test, the children advancing individually as they mastered each exercise.
5. The use of independent extensive reading through the provision of interesting material, each child being permitted to read at his own pleasure.
6. The use of many other accepted remedial measures in reading.

(Bulletins 1, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17.)

Not only was more time spent on test-determined instruction in reading than on any other subject, but much more elaborate planning was done for it. That the results were not as clear-cut as they were in arithmetic may be attributable, in the main, to the complexity of the reading process and to the complexity of the method used in the present study. The work entailed not only much more immediate preparation on the part of the student teacher but involved also problems of teaching technique and class management that might well tax the resources of an inexperienced student teacher.

The final tests, inasmuch as they included exercises in vocabulary meaning, sentence meaning, and paragraph meaning, followed, in method, much the line of the test-determined instruction and since the test-determined instruction further was based upon reading material selected on the basis of difficulty just within the range of the children's varying abilities indicated by their test results, the final tests can be considered valid measures of the test-determined instruction. Since, moreover, the same methods used in this experiment have brought results not only in other experiments but in many public school systems, it seems reasonable to conclude that in the case of reading, the slightness of the advantage of the experimental group over the control is due primarily to the *difficulty of the reading process and of the reading method used when handled by the inexperienced student teacher.* This conclusion is suggestive of the further experimentation

needed along the same general lines as the present study but differing from this in the use of a simpler method of test-determined instruction before definite conclusion can be reached as to the value of teaching students to use the results of tests in reading in their instruction of children.

Arithmetic Problem Solution—The test-determined instruction in arithmetic problem solution consisted of.

1 Rousing desire for improvement through records of achievement, testing, and establishment of definite goals
2 Analysis of errors in problem solution made by the pupils in the Buckingham Scale for Problems in Arithmetic
3 Drill in problem solution based upon class errors with subsequent division of the class into groups for further work as needed
4. Emphasis in instruction upon
a Use of material varied and true to fact
b Thoughtful reading of problems
c Formation of certain clearly defined habits in problem solution, e g, noting what is given in the problem, what is required, and making decision as to the process to be used
d Study of arithmetic data for the purpose of finding possible resulting problems
(Bulletins 6, 7, 8, 10, 11, 14, 17, 19)

The lack of clear-cut results may be attributable to the fact that less time was given to test-determined instruction in problem solution than to any other one of the subjects, the time given to problem solution approximating but one-third of the time given to reading Some of the gain that was made may be attributable to transfer from the test-determined instruction in reading, since, as has been brought out by P. R Stevenson,[1] E. Wilson and others, the ability to solve problems is, to so great an extent, dependent upon reading ability.

Stevenson[2] recommends in addition to the instructional points emphasized above, practice in solving problems without numbers, in dramatizing problems, and definite study of arithmetic vocabulary E Wilson[4] stresses "reading interesting situations into facts" On the other hand, scepticism is expressed as to the value of any arithmetic drill in problem

[1] Stevenson, P R "Difficulties in Problem Solving," Journal of Educational Research, XI, 1925, pp 95-103
[2] Wilson, E "Improving the Ability to Read Arithmetic Problems," Elementary School Journal, XXII, 1922, pp 380-6
[3] Stevenson, *Ibid*
[4] Wilson, *Ibid*

solution. Leta S. Hollingworth, for example, says, "Ability in problem solving can probably not be much affected by drill, since 'a problem' is, by definition, something that requires independent adjustment, and not the response of automatic habit. It, therefore, calls on general intelligence and cannot be improved after the mechanics of reading and calculating have been mastered up to the limits of capacity."[1] Hamilton voices a similar attitude as follows: "Every problem is less of a problem to the extent to which it can be solved by memory of the method by which similar problems have previously been solved. The expression, 'drill in problem-solving,' which some writers have employed is, therefore, a misnomer for, in so far as previous experience in the form of the drill referred to enables an exercise to be solved that exercise is not a problem."[2]

As opposed to the above doubt of the utility of drill in problem solution, R. S. Newcomb supports his thesis that pupils can be taught to solve problems in arithmetic by the following quotation from Thorndike: "Reasoning or selective, inferential thinking is not at all opposed to or independent of the laws of habit, but really is their necessary result under the conditions imposed by man's nature and training. A closer examination of selective thinking will show that no principle beyond the laws of readiness, exercise, and effect are needed to explain it."

In the absence of conclusive experimental evidence as to the improvability of arithmetic problem solution, either by the methods used in this experiment or in any other, the test-determined instruction in problem solution used can neither be supported nor condemned with any degree of certainty. The method outlined in the bulletins for the students was comparatively simple and should not have been nearly so great a tax as the reading method. Since the test-determined instruction was based upon the analysis of the difficulties revealed by the Buckingham Test, the final test might be con-

[1] Hollingsworth, *Ibid.*, p. 1.
[2] Hamilton, E. R. "Insight and Skill in Arithmetic," Journal of Educational Research, XII, 1925, pp. 136-44.

sidered a valid measure of the instruction. In interpreting these results it is important to bear in mind that the time spent on this phase of the work was comparatively small. It is, therefore, concluded *that the slight advantage of the experimental group is indicative of a probable value in teaching students to use the results of tests in problem solution, but that the evidence is not conclusive.*

Summary.—The conclusions derived from the study of the progress of the pupils may be summarized as follows:

1. There is indication that there is some value in teaching students to use the results of general and diagnostic tests in their instruction of children, but the data suggest that there may be varying degrees of value in the different subjects tested: reading, spelling, arithmetic calculation, and arithmetic problem solution.

2. The definite results obtained in arithmetic calculation indicate more value in the use of the method in this subject than in the others.

3. The failure to obtain results in spelling suggests the need for further experimentation in which the emphasis is changed from grouping by type difficulty as in the present experiment to correction of the specific errors revealed in tests and to the cultivation of the habit of learning to spell correctly all words need[illegible].

4. The slight advantage gained in reading as against the comparatively large amount of time used in test-determined instruction in this subject suggest the comparative teaching difficulties encountered by the student in teaching it and indicate the need for devising a simpler technique before putting test-determined instruction in this subject to further experimental test.

5. The results in arithmetic problem solution are suggestive of the value of the method used, but further experimentation along the same lines is needed.

6. The greater degree of correlation between initial status and gains in the experimental group is further evidence of the value of the experimental method, but the lesser correlation between mental age and gains when chronological age is constant is an indication that the method in the hands of the student teacher does not sufficiently favor the superior child.

2. THE SUBORDINATE CRITERION: THE SUCCESS OF THE STUDENT TEACHER

As was suggested in the beginning, the whole matter of teacher rating is too debatable to warrant drawing any conclusions of worth from the results of the present experiment in terms of the final rating of the students. The fact that the average student teaching rating of the city, of the rural, and of the total group was slightly higher in the experimental than in the control is taken merely as very slight additional evidence of the value of the experimental method used.

It may be of some interest to compare the coefficients of correlation obtained between the final student teacher rating and certain other factors with similar coefficients obtained by Knight between general teaching ability of experienced teachers and other factors.[1]

Present Study			Knight's Study	
r Student Tchg. and	Ex.	C.	r General Tchg. and	
Initial rating	.422	.385		
I. E. R. Intelligence	.038	—.035	Intelligence	.164
Jr. Sch. Grades City	.025	.242	Normal Sch. Grades	.147
Jr. Sch. Grades Campus and Rural	.518	.421		
Chr. Age	.289	.144	Chr. Age	.082

While the data are not strictly comparable since Knight's correlations are calculated between general teaching ability and other factors while the correlations in the present study are calculated between student teaching and other factors, they are corroboratory in the case of intelligence, chronological age, and the city Junior Scholarship Grades. The difference between the correlation between city junior scholarship grades and student teaching grades on the one hand and rural junior scholarship grades and student teaching grades on the other is interesting since the scholarship grades for the city students were given in the city training school before it was merged with the normal school and thus represents the ratings of a different group of instructors from those who rated the rural students. The difference suggests the possibility that

[1] Knight, F. B. Qualities Related to Success in Teaching, Teachers College, Columbia University Contributions to Education, No. 120, p. 40.

the more intimate knowledge of the rural students because of their additional year on the campus may have affected the student teaching grades, thus raising the correlation. If so, this might be considered as evidence of the difficulty of making a purely objective rating of teaching, uninfluenced by previous knowledge of the teacher.

While the subordinate criterion adds little, if anything, to the measure of the worth of the method here used, the diaries of the student teachers show evidences of the careful observation of the children resulting from the experimental work. While the data afford no means of measuring such value, it is highly probable that a method resulting in the focussing of the attention of the teacher upon the individual needs of the children will secure improved teaching results.

Summary.—The results judged by the subordinate criterion of the student teaching grades may be considered corroboratory of the results indicated by the main criterion, though subject to the serious limitations of the lack of objectivity in teacher rating, and, hence, not reliable. Evidence in the diary blanks of the students indicative of careful observations of the pupils by the student teachers suggests further value of the method, but the data offer no proof of this point.

CHAPTER XI

Other Problems Suggested by the Data

Five major problems were suggested by the data of the present study. A brief statement of each will be given.

The Administration of Courses in Tests and Measurements.—A problem in this field was suggested by the answers to the questionnaires sent to teacher training institutions (Chapter I). The following procedures in administering courses in tests and measurements were either expressed or implied in the answers:

1. A course in tests and measurements followed by practice in their use during student teaching.
2. The tests and measurements involved in the various subjects taught in the professionalized subject matter courses; e. g., arithmetic measurements in arithmetic courses, reading measurements in reading courses, etc.
3. Tests and measurements included under the study of such topics as intelligence and individual differences in courses in psychology.

If it is granted that the teaching of tests and measurements should be part of the normal school program, it is of importance to know which of these procedures will produce the best results in terms of teaching skill of the students. This problem is suggestive of a large group of problems the solution of which would determine the best of several ways of administering normal school courses.

The Relative Value of Pre-Service and In-Service Training in the Use of Tests and Measurements.—The results of the present study indicate that there is some slight value in teaching student teachers to use the results of tests and measurements. It is possible, however, that such instruction might be more economically given and be more effective if deferred until after the student has entered the actual teaching field. This possibility suggests the problem: Can courses in tests and measurements be more efficiently given during the pre-service or the in-service training of

teachers? Again, this problem is typical of a broad range of research having for its object the determination of the most effective period in the training of teachers for giving instruction in the various elements comprising teacher training preparation.

DIFFERENCE IN INTELLIGENCE STATUS OF THE RURAL AND THE CITY STUDENT.—In equating groups of student teachers according to intelligence, it was noted that the city students ranked considerably higher in the I. E. R. Intelligence Test than the rural students (Figures 6 and 7). These data suggest the need for determining the essential differences between the city and rural students which cause the difference in intelligence rating. To discover these differences should be of practical importance in determining how best to meet the needs of these two groups of students.

THE PROBLEM OF TEACHER RATING.—The difficulty experienced in applying the subordinate criterion, the final rating of the student teachers, suggests that the study of teacher rating might well be made part of the normal school program of research. A normal school offers unusual opportunities for wrestling with this baffling problem since student teachers are, of necessity, under very close supervision of a number of trained observers. Hence, the situation provides to a degree perhaps not obtainable otherwise opportunities for applying the test of objectivity; i. e., that a scale should yield the same results regardless of the individuality of the one using the scale.

THE VALUES RESULTING FROM THE RATING OF TEACHERS.—Another profitable study in this connection might be made of the values resulting from the rating of teachers. It is the opinion of the writer that there is but little such value. Observation over a period of years has indicated that the emotional reaction to a rating seems to outweigh any other value it might have. If a valid criterion can be found for measuring the results in terms of the growth of the pupils taught, a study might well be made of the relative value of the rating as against the non-rating of teachers.

SELECTED BIBLIOGRAPHY

1. The Use of Tests and Measurements in Teacher Training

Grant, Emma B. "Motivating the Course in Tests and Measurements for the Teacher-in-Training." Journal of Educational Method, IV, No. 5 (1925).

Grupe, Mary A., and Smith, Elsa M. "The Use of Educational Measurements in the Training Departments of the State Normal School, Ellensburg, Washington." Educational Administration, VII (1921) pp. 517-26.

James, M. E. "Using the Results of Measurement in Reading in Training Student Teachers." Elementary School Journal, XXIII (1922), pp. 190-96.

Bulletin, 1924, No. 10, Department of the Interior, Bureau of Education, Statistics of Teachers Colleges and Normal Schools, 1921-22, prepared by Frank M. Phillips.

2. The Rating of Teachers

Courtis, S. A. "Standards of Teaching Ability," Educational Review, (LXII), p. 185.

Davidson, W. M. "How to Measure the Efficiency of Teachers," National Education Association Proceedings (1913), pp. 286-290.

Elliott, E. C. "How Shall the Merit of Teachers be Tested and Recorded," Educational Administration and Supervision, I (1915), pp. 291-299.

Knight, F. B. Qualities Related to Success in Teaching, Teachers College, Columbia University Contributions to Education, No. 20, 1922.

Pittinger, B. F. "Problems of Teacher Measurement," Journal of Educational Psychology, VIII, pp. 103-110.

Johnson, F. W. "Supervision of Instruction," School Review, XXX (1922), pp. 742-754.

Kilpatrick, Wm. H. "The Rating of Prospective Teachers," School Review Monograph, No. 5 (1914), p. 12.

3. Statistics

Brooks, Fowler D. "The Reliability of Silent Reading Tests," School and Society, XIX, 492 (1924), p. 652.

Garrett, Henry E. Statistics in Psychology and Education, New York: Longman, Green, and Co., 1922.

Gates, Arthur I. "An Experimental and Statistical Study of Reading and Reading Tests," The Journal of Educational Psychology, XII, No. 7 (21), p. 382.

Illinois Examination I and II Prepared by R. S. Monroe and B. R. Buckingham. Teachers' Handbook. The Public School Publishing Co., Bloomington, Ill., 1920.

McCall, Wm. A. How to Measure in Education, New York: The Macmillan Co., 1922.

McCall, Wm. A. How to Experiment in Education, New York: The Macmillan Co., 1921.

Otis, Arthur S. Statistical Method in Educational Measurement, Chicago: World Book Co., 1925.

Root, W. T. "Correlation between Binet Tests and Group Tests," Journal of Educational Psychology, XIII, No. 5 (1922).

Stanford Achievement Tests by Truman L. Kelley, Giles M. Ruch, and Lewis M. Terman, Manual of Directions, New York: World Book Co., 1924.

THOMSON, GODFREY H. "A Formula to Correct for the Errors of Measurement on the Correlation of Initial Values with Gains," Journal of Experimental Psychology, VII, No. 4 (1924).

THORNDIKE, E. L. Mental and Social Measurements, New York: 1919, Teachers College, Columbia University.

4. READING

ANDERSON, C. J., and MERTON, E. "Remedial Work in Silent Reading," Elementary School Journal, XX (20), pp. 685-701; 772-791.

BURGESS, MAY AYRES. "Classroom Grouping for Silent Reading Drill." Elementary School Journal, XXII (21), pp. 269-78.

BUSWELL, G. T. Fundamental Reading Habits: A Study of Their Development. University of Chicago, 1922.

DOUGHERTY, MARY L. How to Teach Phonics, Boston: Houghton, Mifflin Co., 1923.

GATES, A. L. The Psychology of Reading and Spelling with Special Reference to Disability, New York: 1922, Columbia University.

GRAY, WM. S. Principles of Method in Teaching Reading Derived from Scientific Investigation. Eighteenth Yearbook, of the National Society for the Study of Education, Part II.

GRAY, WM. S. Studies of Elementary School Reading through Standardized Tests. Supplementary Educational Monographs, Vol. 1, No. 1.

HAWLEY, W. E. "Effect of Clear Objectives on the Teaching of Reading," Journal of Educational Research, III, No. 4 (1921), pp. 254-260.

HOLLINGWORTH, L. S. Special Talents and Defects, New York: Macmillan, 1923.

HUNT, C. W. "Extensive Reading; a Factor in Developing Reading Ability," School and Society, XI (20), pp. 260-61.

Maryland School Bulletin, Vol. V, March, 1924, No. 11.

O'BRIEN, J. A. Silent Reading, New York: Macmillan, 1920.

SUTHERLAND, A. H. "Correcting School Disabilities in Reading," Elementary School Journal, XXIII (1922), pp. 37-42.

The Twenty-Fourth Yearbook of the National Society for the Study of Education, Part 1.

ZIRBES, LAURA. "Diagnostic Measures as a Basis for Procedure," Elementary School Journal, XVIII (1918), pp. 505-522.

5. SPELLING

BREED, F. S. "What Words Should Children be Taught to Spell?" Elementary School Journal, XXVI (25) pp. 118-31; 202-14; 292-306.

COURTIS, S. A. Teaching Spelling by Plays and Games, Detroit, 1917.

HOLLINGWORTH, L. S. Special Talents and Defects, New York: Macmillan, 1923.

HORN, ERNEST. Principles of Method in Teaching Spelling as Derived from Scientific Investigation. Eighteenth Yearbook, National Society for the Study of Education, Part II.

KINGSLEY, H. "Test Study Method versus the Study-Test Method in Spelling, Language, and Arithmetic," Elementary School Journal, XXIV (23), pp. 126-9.
REED, HOMER B. The Psychology of Elementary School Subjects, Ginn & Co., 1927, p. 230.
TIDYMAN, W. F. Teaching of Spelling. Yonkers: World Book Co., 1919, p. 16.
GATES, A. L. The Psychology of Reading and Spelling with Special Reference to Disability, New York: Columbia University, 1922.
O'HERN, J. P. "Practical Application of Standard Tests in Spelling, Language, and Arithmetic," Elementary School Journal, XVIII (18), pp. 662-79.

6. ARITHMETIC

Arithmetic Problems Issued by the Division of Reference and Research, Department of Education, New York City.
GRAY, OLIVE. "Teaching Pupils to Read Arithmetic and Other Subject Matter," Elementary School Journal, XXVI, No. 8 (1926).
HAMILTON, E. R. "Insight and Skill in Arithmetic," Journal of Educational Research, XII (26), pp. 136-44.
Maryland School Bulletin, Vol. IV, Oct., 1922, No. 5.
NEWCOMB, R. S. "Teaching Pupils How to Solve Problems in Arithmetic," Elementary School Journal, XXXII (22), pp. 183-89.
STEVENSON, P. R. "Difficulties in Problem Solving," Journal of Educational Research, XI (1925), pp. 95-103.
The Eighteenth Yearbook of the National Society for the Study of Education.
WILSON, E. "Improving the Ability to Read Arithmetic Problems," Elementary School Journal, XXII (22), pp. 380-6.
HOLLINGWORTH, L. S. Special Talents and Defects, New York: Macmillan, 1923.

APPENDIX A

Courses in Educational Measurements in Teacher Training and Normal Schools

I

1. Catalogue name or names of courses in which training in the use of educational tests and measurements is given.....
.....
.....

2. Enrollment:
a. In entire school .....
b. In each course in educational tests and measurements .....
c. Are such courses required or elective? .....

3. Teaching Staff:
a. Number on entire staff of school .....
b. Number teaching courses in educational tests and measurements:
1'. Number devoting entire time to such courses .....
2'. Number giving part time to such courses .....
3'. Where part time is given, state the fraction of entire teaching time given to such instruction .....

4. School Year:
a. Length of school year in weeks .....
b. Length of each course in educational tests and measurements in weeks .....

II

5. Salaries:
a. Total expenditure made upon salaries of the entire teaching staff during 1923-24 .....
b. Total expenditure made upon salaries of instructors giving courses in tests and measurements during 1923-24:
1'. Of those devoting entire time to such instruction .....
2'. Of those giving part time to such instruction .....

6. Expenditures other than salaries:
a. Total expenditure made on equipment and supplies used in conducting courses in educational tests and measurements during 1923-24 .....
b. Total expenditure on assistance other than costs of instruction made during 1923-24 .....

III

7. Testing program in schools outside your own institution:
a. How many schools were so used? .....
b. How many individuals were tested? .....
c. How many different tests were used? .....

8. Cost of testing program:
a. What was the total cost of this testing program (exclusive of all expenditures listed under II.) .....
b. Of this sum, what amount was spent upon testing materials? .....

c. How much of the cost of the testing program was defrayed by your institution? .....
d. What amount, if any, was contributed by other sources? .....
Indicate the sources. .....
9. Was any use other than the most effective training of students made of the data gathered in the testing program? .....
Indicate the uses, if any. .....
Date .....

Signed .....
Office .....

NOTE.—The questionnaire sent to departments and schools of education in colleges and universities was similar to this one, being merely modified in a few details to suit the differences in organization of the two types of institution.

APPENDIX B

COOPERATION OF BUREAUS OF EDUCATIONAL RESEARCH WITH TEACHER TRAINING INSTITUTIONS

I

1. Do you cooperate with any teacher training institution or department or school of education of a college or university? .....
If so, give name and location of such institution. .....
.....

II

2. Does such cooperation include the giving of courses in educational measurements in such institution by members of your staff? .....
3. Give the length of such courses in weeks per year .....
Number periods per week ..... length of periods .....
4. Is the cost of such instruction met:
a. Entirely by the teacher training institution? .....
b. Entirely by the bureau of research? .....
c. By the bureau of research and teacher training institution cooperatively? .....
5. Indicate the amount appropriated for 1923-24 by the bureau of research for such instructional purposes .....

III

6. Are you engaged in any testing program in cooperation with the teacher training institution? .....
7. Estimate after each of the following items the amount spent during 1923-24 in connection with such program:
a. Testing materials .....
b. Equipment for experimental teaching .....
c. Clerical assistance .....
d. Office supplies .....
e. Postage .....
f. Any other item (indicate its character) .....
8. Estimate the total amount spent during 1923-24 upon such cooperative programs .....
9. What benefits, if any, do you consider result from such cooperation?

APPENDIX C

Instructions for Keeping Diary of Test-Determined Instructions

(For Student Teachers in Experimental Classes)

I. The diary blanks which you are asked to keep are essential in evaluating the results of the test-determined instruction in reading, spelling and arithmetic which you are asked to give. Accuracy is essential.

Fill in a diary blank as soon after giving a test-determined lesson as possible.

Keep the diary blanks in a filing folder or in a loose-leaf note book cover available for study by the teacher of practice or the experimenter at any time.

II. Observe the following explanation of items used on the diary blanks:

7. If the time of the lesson is divided between groups, indicate which groups were taught; e. g.—vocabulary group, sentence meaning group, etc. After each group, indicate whether the activity was directed or undirected.

8. Types of instruction.—Use the same terms as used on the sheets of instructions for test-determined reading, arithmetic, and spelling.

9. Materials Used:

a. Give in every case the source of the materials used: text books, books other than text books, magazines, drill devices, other school activities, etc. Indicate the amount of such material used: pages or paragraphs, or number of words, or number of problems according to the material used.

b. If spelling is not taken from text, write the words.

c. If arithmetic is not taken from a text, give a sufficient number of the examples or problems to show the types used.

10. Comments:

Indicate any conditions or occurrences, beneficial or detrimental, that might have influenced the results of the work: e. g.—interruptions, attitude of children.

11. Disciplinary Cases:

Give the name of the child and the nature of the offense of any disciplinary case that occurred which was of such character as to interrupt the lesson.

APPENDIX D

Form 2

DIARY—TEST-DETERMINED INSTRUCTION

1. School 2. Grade 3. Date 4. Time

Activity 1 From .....
To .....
" 2 From .....
To .....
" 3 From .....
To .....

5. No. of Minutes
6. Subject
7. Group or class

8. Types of Instruction:
   Activity 1
   " 2
   " 3
9. Materials Used
10. Comments
11. Disciplinary Cases
12. Absent

Signed _____
Student Teacher

APPENDIX E

SAMPLE OF SCORE SHEET FOR READING TESTS SENT TO EXPERIMENTAL CLASSES WITH BULLETIN 1

October, 1924

EXPERIMENTAL CLASS B

Pupil Num.	Monroe Comp.	Stanford Vocab.	Stanford Sent.	Stanford Par.	Thorndike McCall	Monroe Rate
1	3					
2	4	16	17	14	32	61
3	4	14	6	10	24	206
4	4					
5	6	13	14	22	37	74
6	6					
7	6	11	15	32	31	112
8	6	12	8	26		86
9	7	28	34	48	34	112
10	7	31	28	48		99
11	7	12	24	20	34	125
12	8	26	20	24	40	150
13	8	28	36	18	38	206
14	9	19	21	36	36	150
15	9	16	14	22	34	125
16	9	24	32	36	32	125
17	9				32	136
18	10	24	10	38	34	146
19	10	18	22	24	36	150
20	10	27	30	28	40	150
21	10	27	30	56	40	125
22	10	25	14	18	35	159
23	10	19	10	20	34	146
24	11	28	31	30	49	146
25	11	32	34	28	43	159
26	11					159
27	11	28	33	48	43	146
28	11				43	206
29	12	42	18	60	43	183
30	12	38	26	62	37	206
31	12	28	38	54	38	171
32	12	39	39	60	47	183
33	12	30	28	46	37	159
34	14	40	39	62	41	206
35	14	28	26	46	37	193
Av.	9	24.6	23.8	38.2	36.5	144.8
Grade norms:						
3rd		10.4	12.3	23.4	37.3 (end)	
4th	7.7 (beginning)	26.9	24.1	45.4	39.6 (middle)	122

APPENDIX F

Sample of Suggestions for Grouping in Reading Sent with Bulletin 1

Experimental Class B

N. B. Key numbers for pupils refer to reading score sheets.

REMEDIAL POLICIES IN VOCABULARY

Pupils 2, 3, 5, 7, 8, 11, 15, 16, 18, 19, 23.

REMEDIAL POLICIES IN SENTENCE MEANING

Pupils 2, 3, 5, 7, 8, 11, 12, 14, 15, 18, 22, 23, 29.

REMEDIAL POLICIES IN PARAGRAPH MEANING

Pupils 2, 3, 5, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 25.

PUPILS NOT IN NEED OF REMEDIAL MEASURES

Pupils 27, 30, 31, 32, 33, 34, 35, 21, 24, 26, 28.

NOTES

1. Pupils 3 and 13 are probably reading too fast.
2. Include Pupils 1, 4, 6 (absent for Stanford Test) in vocabulary test.